

Syracuse University

**SURFACE**

---

Theses - ALL

---

6-2014

## Content Externalism and Privileged Access

Kevin Anthony Sadekoski  
*Syracuse University*

Follow this and additional works at: <https://surface.syr.edu/thesis>

---

### Recommended Citation

Sadekoski, Kevin Anthony, "Content Externalism and Privileged Access" (2014). *Theses - ALL*. 33.  
<https://surface.syr.edu/thesis/33>

This is brought to you for free and open access by SURFACE. It has been accepted for inclusion in Theses - ALL by an authorized administrator of SURFACE. For more information, please contact [surface@syr.edu](mailto:surface@syr.edu).

# Abstract

Ever since Descartes, many of us have not questioned the assumption that the knowledge we are able to obtain from accessing our own mental states is a priori in nature since we can access these states outside of experience. Content Externalism has called that assumption into question, it carries the implication that our mental states necessarily depend on particular objects and events that exist in our external environment for at least part of their content. Some philosophers believe that this principle of Privileged Access is inherently incompatible with Content Externalism. Others are convinced that both must be true and compatibility must be possible. This thesis explores arguments from both sides and highlights the merits and disadvantages of each. The paper concludes with the argument that a theory developed by Sarah Sawyer presents the best solution to incompatibility problems.

Content Externalism and Privileged Access  
A Perspective into the Compatibility Debate

by

Kevin Sadekoski

B.A., Syracuse University, 2009  
M.A., Syracuse University, 2014

Thesis

Submitted in partial fulfillment of the requirements for the degree of  
Master of Arts in *Philosophy*.

Syracuse University  
June 2014

Copyright © Kevin Sadekoski 2014  
All Rights Reserved

## Table of Contents

Chapter I – Page 1
Chapter II – Page 5
Chapter III – Page 17
Chapter IV – Page 24
Chapter V – Page 41
Chapter VI – Page 57
Chapter VII – Page 83
Chapter VIII – Page 97
Chapter IX – Page 111
Chapter X – Page 128

# Chapter I

For centuries there has been a Cartesian perspective of the mind which has influenced our beliefs about how we access the content of our own thoughts. Despite the many flaws in Descartes' arguments which led to the development of this Cartesian perspective of the mind, some key features describing our access to our thoughts can be extracted from those arguments which the majority people believe are undeniable. These features are that we are able to access the content of our thoughts without the need for empirical investigation and that no one is in a better position than ourselves to judge whether we are correct in our assessment of that content. The former feature is often described as a priori, or direct, knowledge and the latter is described as authoritative knowledge. While we may not be able to prove beyond a doubt that we have this kind of 'privileged access' to our thoughts, the fact that people seem to access the content of their thoughts in this way on a daily basis is enough to justify our belief in it.

Now this belief in a privileged access to our own thoughts would have become something to be just taken on faith if it wasn't for the fact that it creates a conflict when combined with another theory that the majority of people also believe to be true. This theory is Content Externalism which dictates that at least some of the content of our 'wide' mental states necessarily depend on some relation it bears to the external environment. It is reasonable to believe that any concepts we have which pertain to objects to the external environment will have obtained their content from those objects in the external environment. Therefore it feels safe to assume that the content of any thoughts we have using those concepts must have been derived from the relations the concepts bear to their referents in the external environment. The problem only ensues when we consider the idea that if we are able to know our thoughts and their content

without having to do empirical investigation then it appears that it is possible for us to have a priori knowledge of contingent facts about our external environment. In other words, as long as we assume content externalism is true and we do actually have a privileged access to our own thoughts then seems to be the case that we can have knowledge without empirical investigation which should only be obtainable through empirical investigation.

Many philosophers have developed arguments in attempts to solve the problem while others drew up counter-arguments for why they either did not or could not solve the problem. The philosophers in the former category are called 'Compatibilists' while those in the latter category are called 'Incompatibilists'. The purpose of this thesis is to present the best arguments put forth from both sides and analyze their strengths and weaknesses. The ultimate goal is to figure out if any argument sufficiently navigates around the problems that are created when we combine content externalism and privileged access without sacrificing essential components of either one.

Most Compatibilist arguments usually end up proceeding by assuming a different version of either content externalism or privileged access than was in place when the conflict originated. As long as the version being assumed of either content externalism or privileged access is consistent with our intuitions that led us to believe in their truth in the first place then this would not be an issue. However, most of these alternate versions lack that consistency. The arguments which use a different version of privileged access usually result in only having a priori knowledge to a portion of the content of our thoughts or even just to the fact that a thought has occurred without knowing any of its content. In turn, the arguments which use a different version of content externalism usually loosen the hold that the external environment has on the content of our concepts. The reason these types of arguments are unacceptable is because they don't

solve the same problem that was originally created. Using a different version of content externalism or privileged access to circumvent the issue is basically the same thing as being an Incompatibilist because you are essentially admitting that the two cannot work together as they currently stand so one must be changed.

The second chapter of this thesis begins with an incompatibilist argument by Michael McKinsey. Even though Tyler Burge and Donald Davidson wrote on the subject prior to the development of McKinsey's argument, he deserves to be discussed first because his argument was strong enough to draw the attention of the majority of compatibilist arguments that were developed after the publication of his paper, "Anti-Individualism and Privileged Access"(1991). The following chapter pertains to Anthony Brueckner whose paper, "What an Anti-Individualist knows A Priori"(1992), is a direct response to McKinsey's paper. In Chapter 4 we finally arrive at the views of Donald Davidson and Tyler Burge whose respective papers, "Knowing One's Own Mind"(1987) and "Individualism and Self-Knowledge"(1988), were written prior to McKinsey's paper but present compatibilist arguments which apply to McKinsey's incompatibilist argument. Both author's compatibilist arguments appear strong until we consider Paul Boghossian's arguments in his paper, "Content and Self-Knowledge"(1989). Although this paper was written prior to McKinsey's, Boghossian's arguments in it go into much more detail than McKinsey as to why the kind of self-knowledge that is supposed to result from privileged access is a problem when we assume content externalism. As a result many compatibilist arguments were developed to address the problems presented by Boghossian. Following the chapter on Boghossian we consider such arguments from Kevin Falvey, Joseph Owens, Andre Gallois, and Maria Lasonen-Aarnio. Arguments from each of these authors take a unique perspective on the issue at hand but in one way or another they alter crucial features of either



content externalism or privileged access which are indispensable without changing the debate altogether. For this reason we ultimately end up with arguments present by Sarah Sawyer whose theory takes an unforeseen approach of embracing a conclusion that most people consider absurd. It is through Sawyer that we will find the theory with the best potential for solving this whole problem altogether.

As we proceed through each chapter I discuss each author's arguments as they presented them in their respective papers. After each argument is fully developed it is put up against whatever incompatibilist arguments are the most relevant to see how it handles the problem they create. Lastly I illustrate the weak points of the argument and assess its strength in being a potential solution to the compatibility problem. While every argument has its own merits, it is safe to assume that up until the chapter on Sawyer there will be significant problems with each argument. However, even in insufficient arguments there are sometimes features which are adopted by others and make contributions to the development of new arguments. It is for this reason that many authors were included in this project despite their individual arguments lacking any potential towards being a solution.

## Chapter II

In his paper, “Anti-Individualism and Privileged Access,” Michael McKinsey paints a picture of Privileged Access by first starting with its Cartesian origins and moving forward to the more uncontroversial version commonly accepted today. He defines a priori knowledge as knowledge obtained independently of empirical investigation and Privileged Access as the principle that it is possible to have such knowledge of one’s own neutral cognitive attitude states. (McKinsey 175) McKinsey believes that the intuition that pushes us to accept Privileged Access as an unquestionable principle is just an apparition of the mind. He refers to Tyler Burge’s use of Putnam’s Twin Earth thought experiment in Burge’s 1982 paper “Other Bodies” as reason to believe that two individuals can be exactly the same internally and yet it is possible for them to be in two distinct mental states. Any mental states which could fall into this category are to be considered *wide* states as opposed to *narrow* because part of its content is determined by external factors. McKinsey assumes that due to Burge’s position in “Other Bodies” he would advocate an Anti-Individualist statement like the one below:

- (B) Some neutral cognitive states that are ascribed by *de dicto* attitude sentences (e.g., ‘Oscar is thinking that water is wet’) necessarily depend upon or presuppose the existence of objects external to the person whom the state is ascribed.

Despite the apparent conflict between (B) and Privileged Access, in “Individualism and Self-Knowledge” Burge maintains the Anti-Individualist theory he promotes and Privileged Access are compatible. For the remainder of the paper McKinsey dedicates his effort into showing that both Burge and Donald Davidson, who shares a similar view, are wrong.

It is obvious that the reason McKinsey decides to include a response to Davidson in his paper is that Davidson claims that Incompatibilists are all making a particular mistake which McKinsey does not believe any Incompatibilist has made so far. “Such philosophers make the mistake, Davidson says, of inferring from the fact that a thought is identified or *described* by relating it to something outside the head, that the thought itself must therefore *be* outside the head and hence must be unavailable to privileged access.” (McKinsey 177) McKinsey claims that he has yet to see any evidence that any Incompatibilist has made this mistake and the most convincing reason for advocating Incompatibilism does not assume thoughts to exist outside the head. Although McKinsey does not provide evidence for his own claim that the prominent defenders of Incompatibilism aren’t making this mistake, his argument against Davidson hinges on the fact that the version of Privileged Access Davidson is endorsing is too weak for even most Compatibilists to accept. The crucial move in Davidson’s argument is to distinguish between thoughts and their descriptions so that knowing a thought and knowing its description are two different cases of knowledge. We could have a priori knowledge of a thought whose description can only be known through empirical investigation. However, the description of a thought is really just the content of a thought put into the form of a language. If the only knowledge we have privileged access to is the existence of a thought we are currently having and not to its meaning then this knowledge is very insubstantial. Surely no defender of Compatibilism wants to claim that with an occurrence of a thought we can know a priori that it is occurring but in order to come to know what the thought is about we have to do empirical investigation. Moreover, McKinsey claims that under Davidson’s view we are not even in a privileged position to know that the event is a mental state as opposed to a physical state. (McKinsey 177)

Moving on to McKinsey's assessment of Burge, McKinsey believes that Burge's Compatibility theory does not involve a weak version of Privileged Access like the version he assumes that Davidson is defending. Instead he constructs his triad of propositions which he claims Burge would argue for their consistency due to what he says in his paper, 'Individualism and Self-Knowledge'. The triad, to reiterate from Chapter 1, consists of the following propositions:

- (1) Oscar knows *a priori* that he is thinking that water is wet.
- (2) The proposition that Oscar is thinking that water is wet necessarily depends upon E.
- (3) The proposition E cannot be known *a priori*, but only by empirical investigation.

Again, E is to be interpreted as an 'external proposition' which exhibits the relation between his thought and whatever it is in the external environment that determines the content of the thought. (McKinsey 178) McKinsey openly admits that the consistency of the propositions depends on how we wish to interpret the dependency relation described in the second proposition. He believes that the only reason Burge considers the triad consistent is because Burge interprets the dependency relation as metaphysical necessity.

McKinsey's reason for believing that this is the interpretation Burge uses is extracted from an argument he uses that claims Incompatibilists are making the same mistake Descartes made in his *Meditations*. Descartes believed that he could exist as a mind completely independent from anything physical including physical reality because the existence of himself as well as his thoughts were undeniable but the existence of his body and the physical universe were not. Arnaud argued that Descartes was wrong because the asymmetry in the deniability of the existence of those things can be present while everything that exists is physical in nature

including Descartes mind and thoughts. While it is possible that Descartes could have been right all along, if it is possible that Descartes or even just some his thoughts could not have existed without certain physical objects existing then Descartes deduction of the possibility of existing as a disembodied mind in a non-physical reality is invalid. Furthermore, knowledge that the existence of Descartes or his thoughts depend on something physical in nature is something only obtainable a posteriori. It is this last point that highlights the mistake Burge attributes to Incompatibilists. The fact that the existence of a mental state depends on something physical does not entail that one must know the dependency relation before it is possible to know the mental state. Of all things we can claim to know a priori, the existence of ourselves as a 'thinking being' seems to be the least controversial piece of knowledge to consider as a priori. Although it is up for debate whether my existence depends on the existence of my parents, it is necessarily the case that my existence depends on the fact that an egg was provided the second half of a DNA sequence which resulted in the creation of my DNA sequence and initiated cellular development. Surely is it possible to know a priori that I exist as a 'thinking being' without first having to know the exact physical process that occurred which makes my existence possible. In fact it must be possible since knowledge of that exact physical process is something only knowable a posteriori.

McKinsey agrees that the triad is consistent under his interpretation of Burge's theory. Oscar may very well in fact know a priori he is thinking water is wet while the existence of that thought depends on something physical. The dependency relation is characterized by the proposition E which is only knowable a posteriori but since the dependency is metaphysical entailment it does not need to be known before we can acquire knowledge of the thought. In this case, E would be the proposition 'that water exists'. Oscar does not need to have knowledge of

this to think the thought because for all Oscar knows his thought was about an incorrect statement. If water did not even exist it is still possible for Oscar to have the thought and know it a priori. To deny this is on par with claiming that people who were orphaned as a child and adopted by a family without them ever knowing do not know themselves or that they exist because their existence metaphysically entails that their biological parents existed and they have no knowledge of their biological parents' existence.

However, McKinsey wants to argue that interpreting metaphysical entailment as the dependency relation attributed to Anti-Individualism results in a different version of Anti-Individualism from the one developed by Putnam or even Burge himself. When it is metaphysical entailment that determines whether a given mental state is a wide or narrow state, it necessarily leads to the notion that nearly all of our mental states are wide states. (McKinsey 180) McKinsey turns Burge's own argument against him by first stipulating the premise that it may very well be possible that we could not exist without the existence of the specific pair of our biological parents in the actual world. If metaphysical entailment is the dependency relation that determines wide individuation then for us to be able to be in any mental state whatsoever metaphysically entails the existence of our parents. This entails that every mental state that has ever occurred in a human being is a wide state. Anti-Individualism would at once become both fairly insignificant, because there aren't any narrow states left to distinguish from wide states, and hard to defend, because it instantly becomes a radical theory when it claims that every mental state presupposes the existence of an external object. Radical theories are much easier for opponents to object to since radical theories by nature are controversial.

The consequence of metaphysical entailment brought on Anti-Individualism is the reason McKinsey believes Burge made a mistake when interpreting the dependency relation as

metaphysical entailment. While it solved the problem of the apparent incompatibility between Anti-Individualism and the Principle of Privileged Access, Burge did not foresee that his solution made the concept of a narrow mental state an empty concept. McKinsey thinks that what Putnam and Burge originally had in mind for wide states in Anti-Individualism, at least for Burge before he wrote ‘Individualism and Self-Knowledge’, had to do with the external implications of the content of mental states and not their existence. (McKinsey 181) While the existence of our mental states may depend on the existence of our biological parents, it is not necessarily the case that the content of our mental states do as well. Arguments will always be made that the content of our mental states presuppose their existence so transitively the content also depends on the existence of our parents but arguments like these miss the point. The existence of a mental state may depend on external objects that have no bearing on the meaning of the mental state and the most that knowledge of the dependency relation would tell us is what made the occurrence of the state possible. On the other hand, the content of a mental state is precisely what gives it meaning and knowledge of the dependency relation it has to external objects will not only provide us with knowledge of how the mental state was individuated but also knowledge of the existence of the objects that determined its individuation. According to McKinsey, Putnam and Burge intended the classification of a wide state to, “say something about what it *means* to say that a given person is thinking that water is wet.” (McKinsey 181)

At this point McKinsey introduces his notion of *conceptual implication* as the correct interpretation of the dependency relation. His definition of this notion is that, “a proposition *p* *conceptually implies* a proposition *q* if and only if there is a correct deduction of *q* from *p*, a deduction whose only premises other than *p* are necessary or conceptual truths that are knowable a priori, and each of whose steps follows from previous lines by a self-evident inference rule of

some adequate system of natural deduction.” (McKinsey 181) Disregarding these extremely strict standards, McKinsey refers to Putnam as a reason for the dependency relation to be interpreted as conceptual implication because Putnam defined wide and narrow mental states in terms of *logical* possibility. This combined with the fact that Putnam originally evoked wide and narrow individuation as *Cartesian* concepts leads McKinsey to believe that a narrow state should be defined as, “a state from which the existence of external objects cannot be *deduced*,” and conversely a wide state should be defined as, “[a state] from which the existence of external objects *can* be deduced.” (McKinsey 181) Here McKinsey provides a reinstatement of (B) from the beginning of the paper where the only difference is the ‘necessarily depends upon or presupposes’ is substituted with ‘conceptually implies’. However, when he does this the triad is no longer consistent because the second premise states that Oscar can know E a priori when the third proposition clearly states that E is only knowable a posteriori.

McKinsey considers an objection to his argument before concluding that Anti-Individualism conflicts with Privileged Access when Anti-Individualism is understood properly. He points out another charge Burge makes against Incompatibilists about a mistake some of them may be making. Burge believes some Incompatibilists have been assuming the proposition below:

- (4) Since the proposition that Oscar is thinking that water is wet necessarily depends upon E, no one, including Oscar, could know that Oscar is thinking that water is wet without first knowing E.

If Incompatibilists were making this assumption then Burge would be correct that they were wrong but McKinsey is right in pointing out that this was actually an incorrect assumption on



Burge's part. Anyone who holds (4) to be true is appealing to closure principles which are known to be false. McKinsey's Incompatibility argument does not assume (4) because nowhere is it required that E be known before the thought can be known. Instead his argument is that for it to be possible to know the content of our thought a priori it must be possible to deduce a priori from that knowledge the existence of the physical objects conceptually implicated by the content of the thought. Since knowing the existence of anything physical a priori is an absurd conclusion, either Anti-Individualism contains an inherent problem or the a priori nature of knowledge obtained through Privileged Access is an illusion.

There are a few objections to McKinsey's argument that should be considered before moving on to the Compatibilists and their responses to McKinsey's triad. First off, if we recall the distinction between knowledge of the existence of a mental state and knowledge of the content of a mental state from our discussion of Davidson, perhaps Burge did have metaphysical entailment in mind for wide individuation but only the metaphysical entailment of the content of the mental state and not of the existence of the mental state. In these terms it is not the case that all of our mental states are wide states because the individuation of the states no longer presupposes the conditions that make our own existence possible. The existence of our thought may metaphysically entail the existence of our parents but the content of our thought does not and it would be the metaphysical entailment of the content alone that matters for whether it is determined a wide or narrow state. In this situation, narrow states would be the mental states which do not utilize natural kind concepts because only natural kind concepts metaphysically entail the existence of external objects.

Although there is another objection that can be made disregarding the idea that metaphysical entailment can pertain to the content of a mental state just as conceptual

implication it able to. The strict standard McKinsey sets up for conceptual implication cannot be what Burge and Putnam had in mind for wide individuation either. In its most basic form, for a state to be considered a wide state is simply to say that the content of our thought was determined at least in part by an external object. We would be taking it a step further if we were to say that knowing a priori we are in that state means that we can know the existence of an external object a priori. There are at least three additional pieces of knowledge that must be knowable a priori before the bridge between a priori knowledge of a mental state and the existence of an external object can even begin to be constructed. First of all, we must be able to know a priori that the state we are in is in fact a wide state. The possibility of our ability to acquire even empirical knowledge that our mental state contains at least one natural kind concept is a very controversial subject let alone a priori knowledge of it. Without knowledge that the state is a wide state we can never be warranted in a deduction from knowledge of the state to the existence of external objects. Secondly, we would also need to know a priori the hold that Externalism has on wide mental states, specifically regarding the dependency relation. Even if we can know a priori our mental state is a wide state we need to know that Externalism ensures that if we have a natural kind concept we must have acquired the concept from interacting with the natural kind object in the correct manner. While most philosophers agree that Externalism is a conceptual truth and is knowable a priori, it is not the case that a majority of people who are not philosophers have heard of Externalism and know it well enough to see it as a conceptual truth. In order to know Externalism a priori one must have either encountered the concept at some point in time or have the ability to develop the concept on one's own outside of experience. It is implausible to believe that even people who have this ability have done so because for such a person inexperienced in philosophy an endeavor to take on such a development would take

many, many years. The last piece of knowledge which is required for the deduction is a priori knowledge of a system of logic and its self-evident inference rules. Even though this is the easiest obstacle to overcome in terms of common consensus, it is true that arguments have been made against the validity of the modus ponens inference rule which is considered by many philosophers to be one of the most basic inference rules around today. I am not going to go over this argument because I do think that modus ponens is a valid inference rule, but the crucial point is that it is a required piece of knowledge which is far from being self-evident. These are three pieces of knowledge which are required to move from a priori knowledge of a state to the existence of external objects which seem to always be left out of the discussion especially on the side of the Incompatibilists.

There is also the question to be asked as to whether McKinsey was correct about conceptual implication when claiming that it must involve premises which are either known or only knowable a priori? For instance, the existence of a triangular object conceptually implies that the sum of its angles will be 180 degrees. However, unless someone has taught themselves geometry completely outside of experience, then they will not be able to know the sum of its angles without empirically investigating the object itself or the tenets of geometry. While every human has the inherent capability to teach themselves geometry outside of experience upon encountering the object, it seems unlikely that people with lower IQ's will actually be able to accomplish this feat. So it would seem that there is an important difference between a piece of knowledge being knowable a priori and an individual person having the ability to come to know it a priori. Just because something is knowable a priori for Einstein doesn't mean that everyone has the ability to know it a priori.

Lastly, there is a question that needs to be asked as to whether McKinsey's definition of conceptual implication is proper in the debate about the compatibility of Externalism and Privileged Access? McKinsey's proposed third proposition has it that however we wish to interpret E, it will necessarily have to be something only knowable a posteriori if our ultimate goal is to make the propositions in the McKinsey Argument consistent. However, the way McKinsey defines conceptual implication ensures that whatever E is it will have to be something knowable a priori. So McKinsey will be right that no matter how things are set up, conceptual implication will always lead to the inconsistency of the triad. The only way to remedy this problem is to object to his stipulated definition of conceptual implication. All that would do is call for a whole new argument to be constructed because for any given argument its validity cannot be guaranteed to remain constant when a substitution of definitions for a term contained in the argument occurs. Therefore, McKinsey's argument cannot be required to account for such an objection since it would no longer be the same argument if a different definition for conceptual implication were introduced. Even if McKinsey decides to respond, his response would run along the lines of why the definition they are promoting isn't correct and why his definition is correct rather than actually making any substantial contribution to the compatibility debate.

The trick is that McKinsey set everything up so that the only possible outcome is the outcome he wanted rather than allowing for other possibilities and then proving them wrong. It may very well turn out that McKinsey was right and it was a coincidence that the only apparent possibility was that the triad is inconsistent because in actuality it is the only way things could have been all along. However, if making progress in the debate is our overall goal then without such foreknowledge, the idea of restricting all possibilities to the one McKinsey is trying to

promote does not seem like a smart one. If we are trying to figure out whether E is *only* knowable a priori or a posteriori then conceptual implication should allow for the possibility of both and see which situation most resembles the way things *actually* are. If we are trying to figure out whether McKinsey's stipulated version of conceptual implication is how we should interpret the dependency relation, then we should see what potential outcomes it leads to and which outcome most resembles the way things *actually* are. Instead, McKinsey stipulated both and expects it to be a proven truth that the triad is inconsistent when actually it does not prove anything since the way he set it up does not allow things to happen any other way.

Despite all the problems with the McKinsey argument that I just discussed, one problem remains for the Compatibilist. Even though reaching a priori knowledge of the external world requires a lot of a priori knowledge that is up for debate as to whether we can actually obtain it a priori, the idea that we are *capable* of obtaining a priori knowledge of the external world is a serious problem. Before a Compatibilist can hope to completely defeat the McKinsey argument, they must prove that this capability doesn't exist or that it isn't the problem it appears to be. Since arguing that a priori knowledge of the external world is an unproblematic notion certainly appears to be a futile endeavor, most Compatibilists aim at proving that we cannot acquire, in an a priori fashion, the knowledge of the necessary or conceptual truths required to make the deduction to a priori knowledge of the external world.

## Chapter III

Anthony Brueckner was one of the first to respond to the argument McKinsey presented in his paper, “Anti-Individualism and Privileged Access”. Brueckner starts his paper, “What an Anti-Individualist Knows A Priori”, by pointing out that McKinsey’s argument is based on a incorrect interpretation of the necessary dependence relation involved in Burge’s Anti-Individualism. As you recall from the previous chapters, the triad that McKinsey believes Burge would hold consistent is as follows:

- (1) Oscar knows a priori that he is thinking that water is wet
- (2) The proposition that Oscar is thinking that water is wet necessarily depends upon E
- (3) The proposition E cannot be known a priori, but only by empirical investigation

Brueckner agrees with McKinsey in that the consistency of the triad depends on how we interpret E. According to what McKinsey says in his paper, Brueckner assumes that McKinsey would interpret E as E1 below:

(E1) Oscar inhabits an environment containing H<sub>2</sub>O and not XYZ

Using the only paper where Burge uses the terms ‘necessary dependence’, i.e. “Individualism and Self-Knowledge”, Brueckner takes it that McKinsey’s attribution of (2) to Burge with E being interpreted as E1 comes from a counterfactual proof. If E1 was false and Oscar’s phenomenology, functional structure, behavior, etc., were held fixed, then some of the content of Oscar’s thoughts would have been different from what they actually are. Therefore, the content of the thought depends on E1. As you recall, McKinsey interprets the necessary dependence as conceptual implication which he defines as “a proposition *p* *conceptually implies* a proposition *q* if and only if there is a correct deduction of *q* from *p*, a deduction whose only premises other

than  $p$  are necessary or conceptual truths that are knowable a priori, and each of whose steps follows from previous lines by a self-evident inference rule of some adequate system of natural deduction.” (McKinsey 181)

Brueckner fully concedes that when the necessary dependence in (2) is taken to be conceptual implication under this definition and E is interpreted as E1, the triad will always be inconsistent. However, Brueckner also points out that when (2) is interpreted as such it is not something that an Anti-Individualist advancing Burge’s version of the theory is committed to accept. Under Burge’s Anti-Individualism, one can accept (a), that Oscar would not have been thinking that water is wet had he been in a Twin Earthly environment containing XYZ instead of H<sub>2</sub>O, while rejecting (b), that every world in which Oscar thinks that water is wet is a world containing H<sub>2</sub>O. The proposition (b) is a direct consequence of (2) when E is taken to be E1. (Brueckner 200) The proposition (a) simply implies that in the specific situation we are envisioning Oscar would not have had the thought he did had he been in a different environment when it occurred while (b) is making a sweeping generalization that in any potential situation that has, is, and will ever occur, we could never think a thought about H<sub>2</sub>O in an environment that does not contain it. Even if we weren’t using conceptual implication, E1 has it that Oscar’s ability to think the thought ‘water is wet’ necessarily depends on the fact that Oscar inhabits an environment containing H<sub>2</sub>O.

Brueckner claims that Burge does in fact deny (b) in his paper, “Other Bodies”, which McKinsey alludes to as one of the reasons why conceptual implication should be how we interpret the necessary dependence. Brueckner does not explicate Burge’s denial in detail but he claims that Burge is cautious when discussing conceptual implication because, “such questions concern the possibility of a Kantian transcendental argument against skepticism proceeding from

the assumption of anti-individualism about content.” (Brueckner 200) Essentially Brueckner is claiming that Burge knew that conceptual implication as the necessary dependence inevitably has the consequence that we can access a priori knowledge of the external world. Since knowledge of this kind is impossible to have a priori, Burge knew to tread lightly when discussing the conceptual implication of mental states.

Although this does not completely free Burge’s Anti-Individualism from the criticisms of McKinsey because there is still a way to revive McKinsey’s argument without using E1 or conceptual implication. Brueckner considers the objection that Burge’s thought experiment involving Oscar allows for Oscar to have a priori knowledge of (2c) below:

(2c) If Oscar’s environment had been sufficiently different from the way it in fact is (for example, if it had contained XYZ instead of H<sub>2</sub>O), then, even holding fixed Oscar’s phenomenology, functional structure, behavior, etc., Oscar would not have been thinking that water is wet.

From a priori knowledge of his thought ‘water is wet’ combined with a priori knowledge of (2c) obtained through the thought experiment, Oscar can know a priori that his environment does not contain XYZ. (Brueckner, 201) Brueckner responds to this objection by again claiming that (2c) can be separated into two distinct propositions, (2c1) and (2c2) below:

(2c1) If Oscar’s environment had been sufficiently different from the way it in fact is, then, even holding fixed Oscar’s phenomenology, functional structure, behavior, etc., Oscar would not have been thinking that water is wet.

(2c2) Oscar’s environment in fact contains H<sub>2</sub>O and not XYZ

The former can be known a priori while the latter cannot be known a priori and it is only with a priori knowledge of the latter that we could be said to have a priori knowledge of the external world.

So it seems that even if we interpret E as E1 and use conceptual implication as the necessary dependence, Burge’s Anti-Individualism does inevitably entail the possibility of a



priori knowledge of the external world. The only propositions that a Burgean Anti-Individualist must potentially allow for a priori knowledge of is (2c1) which on its own does not allow for the possibility of accessing a priori knowledge of the external world. Although, objecting to McKinsey's argument was not the only goal of Brueckner's paper because even though it is no longer the case that conceptual implication necessarily leads to the inconsistency of the triad, Brueckner wants to figure out how we should correctly interpret the necessary dependence factor. Brueckner alludes to "Other Bodies" to show that Burge does in fact argue that it is possible for someone to have the concept of water in a world where there is no water or even if water didn't exist on any world. Though, in cases such as this we need reason to believe that it is in fact a water concept rather than a twater concept or any other water counterfeit.

There are two possible ways through which someone can have a water concept in a waterless world. The first is by acquiring the concept from a linguistic community which obtained the concept through some members of the community theorizing about a liquid with the chemical makeup  $H_2O$ . The other possible way is that the person acquires the concept by theorizing about a liquid with the chemical makeup  $H_2O$  on their own, which accounts for the possibility of having a water concept in a waterless world when the individual is alone in the world. In both cases it is assumed that enough physical entities exist in the world so that whoever is doing the theorizing can distinguish the concepts of those entities from their theoretical water concept. (Brueckner 202) Through Brueckner's interpretation of Burge's claim about the possibility of having a water concept in a waterless world, Brueckner was able to postulate (N) below as the true characterization of the necessary dependence in (2):

(N) It is necessary that if Oscar is thinking that water is wet, then either (i) water exists, or (ii) Oscar theorizes that  $H_2O$  exists, or (iii) Oscar is part of a community of speakers some of whom theorize that  $H_2O$  exists.

Through this characterization we are now in a better position to make an attempt at deducing how we should interpret the necessary dependence relation that ensures either (i), (ii), or (iii) must be the case if Oscar is able to have the thought in question. Is (N) true in virtue of conceptual implication, metaphysical entailment, or something else? While Burge never answers this question directly, Brueckner claims that Burge does not state once in the paper from which (N) was extracted that the necessity he had in mind was conceptual and knowable a priori. Brueckner also believes that since (ii) and (iii) pertain to how chemical theory reveals the nature of water, Burge had intended (N) to be taken as a metaphysical necessity which is sometimes only knowable a posteriori. Since (N) is a metaphysical necessity, Brueckner argues that Burge's Anti-Individualism does not require that (N) is knowable a priori despite the fact that it is metaphysically entailed from a thought that is known a priori. (Brueckner 203)

Though Brueckner's job is not done because there is an objection to his argument through raising the question of whether Anti-Individualism or even just some parts of the theory can be known a priori. For instance, even if (N) is not knowable a priori, it may be the case that the proposition (P) below can be known a priori:

(P) It is necessary that if Oscar is thinking that water is wet, there exist some physical entities distinct from Oscar.

While Brueckner suspends judgment as to whether (P) is knowable a priori, he admits that if it is knowable a priori then this knowledge combined with a priori knowledge of a thought would allow for a priori knowledge that the world, "contains physical entities sufficient to fix the contents of his thoughts." (Brueckner 204) The possibility of having a priori knowledge of (P) is an extremely controversial claim, but Brueckner believes that it even if it is possible it will not be a problem for Compatibilism. Combining a priori knowledge of (P) and the thought described in

it would at most grant us a priori knowledge that some physical entity exists which is distinct from us. McKinsey believes such knowledge is impossible but Brueckner thinks differently. While such knowledge is implausible, Brueckner believes that to say it is impossible is to make the claim that there could never be a successful Kantian transcendental argument. It is not the case that just because so far no adequate Kantian transcendental argument has been offered the existence of such an argument is impossible. It is certainly possible that Transcendentalism is true but it is an extremely hard theory to make a case for its truth. If McKinsey wants his claim, that a priori knowledge of the existence of some physical entity distinct from oneself, to have any traction he must provide a full-fledged rejection of Transcendentalism as a whole.

Even though Brueckner has backed McKinsey into a corner where the only way out is through the complete rejection of Transcendentalism, his argument does not help the compatibility cause very much. While Brueckner's argument shows how difficult it would be to make the move from a priori knowledge of our thought to a priori knowledge that a given entity exists in our external environment, sacrifices were made along the way that force us to diminish our intuitions about privileged access. When we have a thought about water, we know it is about water and not some alternative concept because we know how we acquired our water concept and the substance that our water concept refers to. We may not know that we actually interacted with the substance as a natural kind in our external environment for the same reasons that we cannot know whether the external environment we perceive through our senses is even real. However, as long as we proceed on the assumption that in our perceptions of the external environment we correctly perceive existing objects then we do actually know that our concept of water used in our thought is of entities that exist in the external environment. Any compatibilist

theory must be able to account for this if it is to be considered a potential solution to the problem so Brueckner's theory does not make the cut.

## Chapter IV

Donald Davidson is considered by many to play a very significant role in the compatibility debate between Privileged Access and Anti-Individualism (Content Externalism). While he does have much to say on the subject in general, his biggest contributions to this topic are a few key ideas and thought experiments presented in his paper, “Knowing One’s Own Mind”, as well as his overall argument for compatibility. These ideas were eventually adopted and used by other authors who proceeded to make big steps toward a Compatibilist solution. The most significant of Davidson’s ideas is without a doubt his introduction of first- and second-order mental states. Many authors, including Tyler Burge, Kevin Falvey, and Joseph Owens, have used the concepts of first- and second-order mental states directly in their proposed solutions despite the fact that Davidson never explicitly uses those terms himself. Later on in this chapter we will discuss Tyler Burge’s Basic Self-Knowledge which uses the nature of second-order mental states to develop a set of self-verifying self-knowledge. As we will see when we get to the chapter on Paul Boghossian, Davidson’s own argument for compatibility doesn’t come close to addressing all the problems associated with the apparent incompatibility of Privileged Access and Anti-Individualism. However, his ideas did help others develop arguments which have addressed many more of the problems that Davidson’s argument missed and have even been considered by some to have solved the whole problem altogether.

Before getting into Davidson’s ideas, thought experiments, and arguments, there is one other significant contribution that Davidson made to the debate and that is providing one of the clearest illustrations of the initial compatibility problem. It all comes down to the fact that our method of determining what other people think is by gathering evidence based on their behavior

and our knowledge of the person. However, our methods for determining what is going on in our own mind rarely involve the use of evidence even when it is readily available. "It is seldom the case that I need or appeal to evidence or observation in order to find out what I believe; normally I know what I think before I speak or act. Even when I have evidence, I seldom make use of it." (Davidson 87) Perhaps all that means is that most people have been conditioned by the external world to give the concept of evidence more weight than it deserves. However, what is more likely is that it should cause us to question the way we actually do know our own minds.

Disregarding a drastic paradigm-changing discovery such as finding out what we think we know of our own minds is just an illusion, we can be extremely sure that we do know our own minds. Whether or not our knowledge of our own minds is a priori or infallible is still up for questioning, but the fact that when we have a thought we know we are having that thought seems pretty undeniable. Furthermore, we can come to know what we are thinking in the absence of or prior to any evidence being available. On the other hand, to even be able to attempt to know the minds of others we must use evidence and without any evidence to use we are completely unable to make any determination. All the while it is almost universally accepted that the knowledge of our own minds determined without evidence is more reliable than the knowledge of other people's minds determined with evidence. "It is a strange idea that claims made without evidential or observational support should be favored over claims with such support." (Davidson 88) It is important to note that Davidson believes we can be mistaken about our own thoughts so we are not infallible when it comes to knowing our own minds. However, even this addition does not change our intuition that our knowledge of our own minds which usually lacks evidential support seems more reliable than our knowledge of the minds of others which requires evidential support.

One unique point made by Davidson is that if/when we are ever wrong about what is in our own minds then it cannot be that we were led astray by incorrect evidence. Although, he doesn't address the fact that we probably could claim that the error was at least in part due to a lack of evidence. Regardless, we tend to be correct about our own minds with a lack of evidence and there seems to be no reason why we should trust claims made without evidence more than we trust claims made with evidence. According to Davidson, this all comes down to a question; "Setting aside, then, self-deception and other anomalous or borderline phenomena, the question is whether we can, without irrationality, inconsistency, or confusion, simply and straightforwardly think we have a belief we do not have, or think we do not have a belief we do have." (Davidson 90) This quote will also be important later because it is one of the two places where Davidson's concept of second-order mental states can be extracted.

Hilary Putnam presented what seems to be the biggest obstacle to overcome in answering this question with his Twin Earth thought experiment. By this point we know Twin Earth all too well so we know that Putnam argued that what words mean depend on more than 'what is in the head' and therefore the meaning of our thoughts using those words will depend on the same factors which cannot be found within us. Davidson offers the best description of the problem at hand:

"Putnam holds that many philosophers have wrongly assumed that psychological states like belief and knowing the meaning of a word are both (I) 'inner' in the sense that they do not presuppose the existence of any individual other than the subject to whom the state is ascribed, and (II) that these are the very states which we normally identify and individuate as we do beliefs and the other propositional attitudes. Since we normally identify and individuate mental states and meaning in terms partly of the relations to

objects and events other than the subject, Putnam believes (I) and (II) come apart: In his opinion, no states can satisfy both conditions.” (Davidson 91-92)

Davidson states here that the ultimate purpose of his paper is to show that ordinary mental states can and often do satisfy both conditions (I) and (II). For Davidson, a mental state of belief can be ‘inner’ by being identical to a state of the body which occurs within the body and thus be identified without referencing anything outside the body. At the same time this belief has a meaning through its content which is ‘nonindividualistic’ due to the fact that at least part of its meaning was derived from the external environment and thus identified in part by its relations to things within the external environment. (Davidson 92)

This brings us to the sunburn thought experiment which is one of the significant contributions that Davidson has made to the debate. One of the reasons that Burge doesn’t think an ordinary mental state can satisfy both (I) and (II) is because of Putnam’s distinction between ‘narrow’ and ‘wide’ psychological states. According to Putnam, narrow psychological states are mental states which do not refer to anything in the external environment while wide psychological states are those that do. Davidson believes that Putnam’s reason for making such a distinction rests on two “largely unquestioned assumptions”:

- (1) If a thought is identified by a relation to something outside the head, it isn’t wholly in the head.
- (2) If a thought isn’t wholly in the head, it can’t be ‘grasped’ by the mind in the way required by the first person authority. (Davidson 102)

Since ‘narrow’ psychological states are a very small set of all the mental states to which our intuitions tell us that we have privileged access, Davidson wants to object to (1). “It should be clear that it doesn’t follow, simply from the fact that meanings are identified in part by relations to objects outside the head, that meanings aren’t in the head. To suppose this would be as bad as to argue that because my being sunburned presupposes the existence of the sun, my sunburn isn’t



a condition of my skin.” (Davidson 103) If we were to take two people with physically identical skin conditions where one was caused by sunburn and the other by some other means then one presupposes the existence of an external object while the other either doesn’t presuppose the existence of any external object or at the very least it presupposes the existence of a different object. This shows that it’s possible for two people to be in physically identical states despite the causes and meanings behind those states being different. For this reason Davidson argues that we should view all psychological states in the same way. In the hypothetical scenario where two people have brain states which are physically identical, there is no reason to think that the content of the mental states associated to their respective brain states can’t refer to distinct objects in the external environment. “Individual states and events don’t *conceptually* presuppose anything in themselves; some of their *descriptions* may, however.” (Davidson 103)

The sunburn thought experiment does provide an argument for how (I) and (II) can both be satisfied by an ordinary mental state. As long as ‘inner’ in (I) is defined as being identical to a state of the body and not identified by referring to objects or events outside the body then the fact that they don’t conceptually presuppose the existence of external objects allows them to satisfy (I). In the sunburn experiment even though the causes of each person’s skin conditions were different, the conditions in themselves were physically identical and both identifiable without reference to anything outside of the body. However, once we focus on the descriptions of the skin conditions, which would include their causal histories, then we find that the description of at least one of them presupposes an external object. For ordinary mental states, to describe them is essentially to state their meaning or content which is why we usually feel like our mental states in themselves are identified in part by the relations to external objects which give them their content. The states themselves do not presuppose anything but it is hard for us to imagine our

mental states without their content because we seem to always know the content of our mental states the instant they occur. So ordinary mental states when considered separately from their content satisfy (I) but the fact that they have content which almost always have relations to external objects or events ensures that they satisfy (II).

Even though Davidson has at this point accomplished what he stated was the purpose of paper, he still hasn't explained how we can have privileged access to the content of our mental states and not just to the fact that they have occurred within us. This is what Davidson's compatibility argument is meant to solve. Throughout the paper Davidson makes claims that we shouldn't view the content of our mental states as 'objects' which are supposed to be 'before the mind' or 'grasped' by it. (Davidson 107) He brings back his sunburn thought experiment by explaining that a 'sunburn expert' who inspects the two people's skin conditions could not, through inspection alone, determine which is the actual sunburn and which is the one caused by other means. By analogy, without empirical investigation we can know that a mental state is occurring within us but we cannot know the description or meaning of the state. According to Davidson, when we view the content of our mental states as objects which play a role in the individuation of our mental states we cannot know the identity of the states without knowing the object that is its content. In turn, we cannot know the object without knowing what determines the identity of the object. "For if it to be in a state of mind is for the mind to be in some relation like grasping to an object, then whatever helps determine what object is it must equally be grasped if the mind is to know what state it is in." (Davidson 106) As long as we view the content of our mental states as objects in this way then it seems impossible that we could ever know the content of our mental states. "For unless one knows *everything* about the object, there will always be senses in which one does not know what object it is" (Davidson 108)

Davidson's solution for this problem is that we stop viewing the content as objects and instead we should identify them in part by their causal history. The way in which we learned our concepts and what has prompted us to apply them is what identifies our mental states that use them. It is not the case that in order to know the content of our thoughts that we must know everything there is to know about the concepts we utilize in those thoughts because that would include knowing how people in our social environment use the word and the referents of those concepts in the external world. In other words, when we do not see the content of our mental states as objects it is no longer the case that we must know everything about the concepts which play a role in the content before we can be said to have knowledge of the content.

Davidson's compatibility argument does offer a lot to the compatibilist position; as we will see later on it does resemble Sarah Sawyer's compatibility argument which I believe does actually solve the problem altogether. However, Davidson's argument has the consequence that virtually no one's concepts could be the same because no two people would have the same causal history for their concepts even if they carry the same label. The reason we commonly thought of the content of our mental states as objects was because we wanted to know that when two people used the same word that they are referring to the same concept. This by no means proves that Davidson's argument is wrong because it may very well turn out that we think we understand each other when in reality everyone just talks to each other in their own unique languages which happen to have similar words. However, if Davidson is right then without further development of his argument to account for this then we are really just substituting one problem for another.

The biggest contribution that Davidson has made to this debate is something that he never explicitly states but other authors do and use it in their theories directly. This contribution is the idea of first- and second-order mental states. There are two quotes in this paper form which this

idea can be extracted. “Setting aside, then, self-deception and other anomalous or borderline phenomena, the question is whether we can, without irrationality, inconsistency, or confusion, simply and straightforwardly think we have a belief we do not have, or think we do not have a belief we do have.” (Davidson 90) “In general, the belief that one has a thought is enough to justify that belief.” (Davidson 88) As we will see when discussing Tyler Burge, his postulation of ‘Basic Self-Knowledge’ seems to be directly drawn from the latter quote. Essentially for Burge, Basic Self-Knowledge is a second-order propositional attitude like judgment or belief whose object is a first-order thought occurring simultaneously. Burge believes Basic Self-Knowledge cannot be wrong because the occurrence of a second-order judgment like ‘I judge that I am thinking water is wet’ is self-verifying. To judge or believe that we have a thought is enough to make the judgment true because making the judgment or forming the belief ensure that we are having the thought.

Up until this point no one had been discussing different levels of thought being involved in the process of how we come to know the content of our thoughts. Davidson himself never actually wrote about it explicitly either. Instead, he just focused on thoughts and beliefs which happened to fall into the category that was later titled ‘second-order mental states’. However, years after Davidson wrote this paper Kevin Falvey and Joseph Owens wrote a paper titled, ‘Externalism, Self-Knowledge, and Skepticism’, which is the topic of a later chapter. In that paper, Falvey and Owens develop an argument for why we can have introspective knowledge to the content of our thoughts based on the nature of second-order states. Their argument is that since the concepts of our first-order mental states are derived from the external environment and our second-order mental states draw their concepts directly from the first-order states, some concepts we would normally consider to be relevant alternatives to other concepts are not

actually relevant because we could not have the former concepts in certain environments. This will be explained more in the chapter on Falvey and Owens but the crucial point is that the whole concept of second-order mental states plays an essential role in arguments being developed many years after Davidson wrote his paper.

When it comes to the compatibility debate between Anti-Individualism and Privileged Access, Donald Davidson is like the Bob Dylan to Rock and Roll. He was a major influence to many that followed him, especially to the compatibilists, but he never received the credit that he deserved. Davidson's overall argument came the closest to an actual solution to the whole problem but the way he presented it didn't make it very appealing to readers because it was vague, a bit confusing, and highlighted its weaknesses rather than obscuring them. Despite the errors, however, his influence can be seen all the way to Sarah Sawyer which is, in my opinion, where the road of this debate ends. So instead of focusing on how much progress Davidson has lent to the compatibility cause himself, we should focus on his ideas that others adopted and made progress in the cause through their development.

In his paper, "Individualism and the Mental", Tyler Burge developed an anti-individualistic theory concerning the individuation of mental states which, through some interpretations of the theory, had serious implications pertaining to the potential knowledge we can acquire about our mental states. We can assume that Burge was unaware of these implications while writing the paper because a few years later he began writing "Individualism and Self-Knowledge" as an attempt to explicate the problem generated by his theory and offer a potential solution to the problem.

Burge states that the problem is created when one combines his Anti-Individualism regarding the individuation of mental states with a restricted Cartesian perspective of Self-

Knowledge, which essentially boils down to a version of the Principle of Privileged Access.

Despite all the problems with Descartes's views of Self-Knowledge, there are two features we can trace to him which seem to be undeniable. These two features are the directness and certainty in which we come to know our own mental states compared to how we come to know the mental states of others. When some of the content of our mental states is determined by the external environment, it is difficult to see how we can have the ability to know our mental states and their content with the directness and certainty that we appear to have without first investigating the external environment. Although Burge's main goal in his paper is neither to explain nor defend his anti-individualism or the restricted Cartesian conception of Self-Knowledge. (Burge 111) Instead he wants to make the problem that occurs when the two are combined visible and then offer a class of Self-Knowledge, which he labels 'Basic Self-Knowledge', with certain characteristics that allow us to know them with the directness and certainty required by the restricted Cartesian perspective of Self-Knowledge without conflicting with his theory of Anti-Individualism. While basic self-knowledge does not completely solve the problem because it only accounts for a small sub-set of all mental states, it identifies a class of self-knowledge which must be true and thus knowable a priori. It puts the ball in the court of the Incompatibilist to explain how we can't know them a priori because it leads to a priori knowledge of the external world.

Burge derives his prime examples of basic self-knowledge from Descartes's famous cogito argument. The force of the argument is drawn from the idea that my ability to think that I exist allows me to know that I do exist because I would not be able to think about my existence, or think about anything for that matter, if I did not actually exist. Since our own existence is

necessary for us to have the ability to think about our existence, any occurrence of a thought ‘I exist’ is necessarily true and thus knowable a priori.

Burge takes this paradigm and applies it to our propositional attitudes. Two examples that he gives are, ‘I think (with this very thought) that writing requires concentration’ and ‘I judge (or doubt) that water is more common than mercury’. (Burge 111-112) The propositions ‘writing requires concentration’ and ‘water is more common than mercury’ may or may not be true independent of a person’s attitude towards them. However, when a person is thinking the proposition ‘water is more common than mercury’ and then develops the propositional attitude ‘I judge that water is more common than mercury’, the propositional attitude must be true because the act of making the judgment ensures its truth. The only way that the propositional attitude wouldn’t be true is if the person’s mental state characterized by that exact propositional attitude did not occur. Just like Descartes’s cogito, the existence of the mental state pertaining to the judgment ensures that the judgment itself is true. The same goes for thoughts about the proposition ‘writing requires concentration’. The proposition itself may not be true because it is possible that some people have the ability to write without concentrating. Regardless, when someone has the thought ‘I think (with this very thought) that writing requires concentration’ the truth of the individual proposition ‘writing requires concentration’ makes no difference to the truth of the thought because the act of thinking the thought alone ensures that it is true.

Later on in his paper, Burge offers a more detailed description of basic self-knowledge in terms of first and second-order mental states which he borrows from Davidson. Essentially basic self-knowledge takes the form of a second-order judgment whose content is a first-order thought. While having a first-order thought we, at the very same time, think about the thought as our own. According to Anti-Individualism, the content of the first-order thought is derived from the

external environment and locked into the conditions from which it was acquired. As an example we can take the commonly used thought 'Water is wet' to represent the first-order thought. Second-order judgments have a reflexive and self-referential nature in that upon having the thought 'Water is wet' we simultaneously judge in ourselves that 'I am thinking that water is wet'. (Burge 122) This second-order judgment is self-referential because we are truly just thinking the same first-order thought occurring within us but we think of it as our own thought and not just as concepts acquired from the external environment that we put together in a way that represents that environment. The second-order judgment is also reflexive because the very act of having the first-order thought naturally causes us to *think* that we are having that first-order thought and therefore judge ourselves to be thinking it. Since the content of the first-order thought is contained within itself, that content is transferred to the content of the second-order judgment because the object of reference for the second-order judgment is the first-order thought itself. (Burge 121)

It is worth pointing out that some would argue that basic self-knowledge can only avoid the problem of being a priori knowledge about contingent facts of the external environment if the content contained in the first-order thought is closed off from our awareness at the second-order level. This is something Burge does not address because his purpose with basic self-knowledge is not to definitively solve the compatibility problem between Privileged Access and Anti-Individualism, but rather to label a class of knowledge which could be compatible and leave it to Incompatibilists to say why it isn't compatible. This objection would be one such response by Incompatibilists but it still remains to be seen if it has any real traction in the debate. One potential counter-response is that perhaps we are only aware at the second-order level of our *perception* of the content of the first-order thought. While we may perceive the content correctly,



there is no guarantee that we are and therefore no guarantee that it actually represents the external environment. This begs the question of whether we could truly consider it knowledge if the perception may or may not be accurate, but this is something Burge actually does address when he compares what we commonly count as perceptual knowledge to what we should consider self-knowledge. At the very least, this introduction of first and second-order mental states to basic self-knowledge helps avoid potential counterfeit objections. The first-order thought will always be the object of reference for the second-order judgment. “Basic self-knowledge is self-referential in a way that insures that the object of reference just is the thought being thought. If background conditions are different enough so that there is another object of reference in one’s self-referential thinking, they are also different enough so that there is another thought.” (Burge 121-122)

As Burge states, “it is certainly plausible that these sorts of judgments or thoughts constitute knowledge, that they are not products of ordinary empirical investigation, and that they are peculiarly direct and authoritative.” (Burge 112) As long as mental states of this kind count as knowledge and didn’t result from empirical investigation then they can meet the ‘knowable a priori’ feature of Privileged Access because they are self-verifying due to the fact that their existence alone ensures their truth. We can know them directly because as long as they exist they are necessarily true and for the same reason we know them with certainty.

Since it is clear that mental states of this kind are self-verifying and are not the result of empirical investigation, all that is left to do so that they may be classified under the label ‘Basic Self-Knowledge’ is to prove that they count as knowledge and not just trivial bits of information. This is a daunting task which many would believe impossible to accomplish since there is a whole field of philosophy devoted to figuring out the conditions that must be met for something

to be considered knowledge. Burge takes on this epistemic endeavor through assessing the conditions we tend to require by common practice for something to count as perceptual knowledge. In doing so he compares and contrasts perceptual knowledge and self-knowledge so that we may see which conditions should be kept due to their similarities and which should be revised or discarded due to their differences. Due to skepticism it is rare in philosophy that common practice can be used to prove anything. To address this Burge states, “My discussion of knowledge and individualism has proceeded on the unargued assumption that skepticism is mistaken.” (Burge 117) With this statement alone he allows for common practice to actually have some traction in his argument because it cannot be defeated by the claims that one person thinks otherwise or that the entire population who believe in the common practice could be wrong.

The first aspect of perceptual knowledge that Burge thinks applies equally to self-knowledge is that in order to have perceptual knowledge of something we are not required to first have knowledge of the conditions that made the perceptual knowledge possible. (Burge 117) To grant someone the perceptual knowledge ‘there is a pool over there’, we do not require that they first must know how their sense of sight works, that they are not hallucinating, that the object is not a counterfeit like a pond made to look like a pool, etc. All that we require for it to be considered knowledge is that they perceived an object that meets the conditions to be classified under their concept of a pool. Likewise, for us to classify something as self-knowledge we should not require that we know all of the conditions that made our acquisition of the knowledge possible. For instance, when we have the thought ‘I think (with this very thought) that writing requires concentration’ we should not be required to know the concepts of writing, requirement, or concentration before we can know the occurrence of the thought as well as which terms are

used in its content. It is true that knowledge of the concepts would allow for a richer and more thorough understanding of the thought which would lead one to believe that the knowledge of the thought is richer and more thorough as well. In the same regard knowledge of how sight works, that you are not hallucinating, or seeing a counterfeit would seem to add more depth to our perceptual knowledge than if it lacked knowledge those things. However, as long as skepticism is assumed to be mistaken and the conditions for what counts as knowledge in common practice are kept the same, then just because a 'deeper' knowledge can be found doesn't mean that the more 'shallow' knowledge shouldn't actually count as knowledge.

Despite the similarity between perceptual knowledge and self-knowledge described above, there is a big difference which leads Burge to believe that it is less plausible than with perceptual knowledge to think that in order to consider basic self-knowledge as knowledge that we need to be able to differentiate the thoughts involved in such knowledge from 'twin thoughts'. This is one form of counterfeit objections like I described in the discussion of first and second-order mental states. Burge's reason for this claim is that perceptual knowledge has forms of objectivity that basic self-knowledge does not share. (Burge 119) The first form of objectivity in perception deals with the possibility of error. We have no guarantee that our perception of objects actually matches the true nature of objects outside of that perception. This could be due to any number of things like illusion, hallucination, or just the brute fact that the process of perception may adjust the true nature of things in the same way that light appears to bend when going through different mediums. This is not possible in basic self-knowledge because it is self-referential and self-verifying. (Burge 120) The second-order judgment must be true because if the first-order thought were different enough to have any effect on the truth of the second-order

judgment then we would have a different second-order judgment altogether instead of a false one.

The other form of objectivity in perception is that we can compare our own perceptions of an object to other people's perceptions of the same object. Science would not be our method of determining truth in our physical world without our ability to recreate experiments that have already been done so that we can test for errors and variations. With perception we have no way of knowing if you and I perceive the same things when we look at the same object without being able to compare our perceptions. With basic self-knowledge there cannot be this kind of objectivity because it must be done from the first-person point of view. (Burge 121) Basic self-knowledge is self-verifying because the first-order thought and the second-order judgment occur at the same time in the same individual in such a way that the second-order judgment cannot be false as long as the first-order thought is also occurring. There is no way that another person could be witness to this process in any way for them to be able to offer any further information that would prevent basic self-knowledge from being self-verifying. Even if an invention were created that allowed us to see the thoughts of other individuals, all they would see is that we had a thought and that we thought about it as our own in such a way that the second-order judgment must be true regardless of the content of the first-order thought. We could have the first-order thought ' $1+1=3$ ' and then the second-order judgment 'I am thinking that  $1+1=3$ ' and it still couldn't be wrong because the truth of the first-order thought does not change the fact that we are thinking it.

While Burge's basic self-knowledge is only a small set of all the self-knowledge that must be made compatible with Privileged Access and Anti-Individualism, its introduction does offer a good stepping stone on the way to compatibility. Its weakest point is whether or not such

second-order judgments can actually be classified as knowledge. However, with skepticism on the table that is a problem for practically everything we have ever wanted to consider knowledge. All in all Burge has given Compatibilists hope that finding a compatibility between Privileged Access and Anti-Individualism is truly possible.

## Chapter V

In his paper, 'Content and Self-Knowledge,' Paul Boghossian starts off by explaining the overall purpose of his paper. The purpose is, "that, given a certain apparently inevitable thesis about content, we could not know our own minds. The thesis is that the content of a thought is determined by its relational properties." (Boghossian 149) I would like to assume here that Boghossian meant *part* of the content of a thought is determined by its relational properties but I am not so sure I can warrant that assumption. Some of Boghossian's main arguments hinges on the idea that there are no intrinsic properties of a mental event, i.e. an occurring mental state, since even the internal factors that determine the content of the mental event are relational properties of the event itself which he claims is the 'currently prevailing orthodoxy'. (Boghossian 161) If this is true then every property of a mental state is an extrinsic property. We will look at this a bit later because it is a controversial point and it lies at the foundation of Boghossian's arguments.

Boghossian proceeds by explaining that the problem is that occasionally when we know our own thoughts we seem to do so in a direct manner without having to infer this knowledge from anything else. If we don't use inference to obtain this knowledge then the only other options are that we either use some form of inner observation or there is no process to describe how we acquired the knowledge, we know our thought on the basis of nothing. However, Boghossian believes that neither of the two options are an adequate description of how we come to know our own thoughts. Both options give us reason to doubt that they will yield the kind of knowledge of our mental states that we occasionally possess. The first section of his paper is dedicated to showing that inference is not the manner in which we can come to know our own

thoughts. Despite the fact that it is already stipulated that we do not use inference for this knowledge Boghossian must have felt that it was useful for his argument to show why it was stipulated in the first place. The second section attempts to show that inner observation will never give us knowledge of a mental state because the properties that we are supposed to observe cannot be perceived through observation alone. In the last section, Boghossian argues that we do not know our thoughts, at least the majority of them, on the basis of nothing. Even if we gain something from our thoughts on the basis of nothing is it not considered knowledge the way he defines it. He states that in his use of the term ‘knowledge’, particularly ‘self-knowledge’, it must be a true and justified belief about one’s own thoughts. (Boghossian 150) As long as proper evidence is required for justification Boghossian finds it hard to believe that any ‘knowledge’ based on inner observation or nothing will be able to provide it.

The first section starts by referencing the perspective of the mind and self-knowledge developed by Descartes which has greatly influenced the Philosophy of Mind for centuries. Despite the obvious flaws in the Cartesian picture of the mind which have been rejected by the majority of philosophers today, some aspects contain a truth about the nature of the mind that cannot be ignored or denied. Even though we are surely fallible when it comes to our judgments about our mental states, there does seem to be an asymmetry in the way we come to know our own mental states compared to how we come to know the mental states of others. No matter how well you know someone else the only way we are able to make a judgment about what mental state they are, were, or will be in is if we observe their behavior and make an inference. When it comes to knowing my own mental state not only is inference from observing my own behavior hardly necessary but it runs counter to the intuitions about the direct nature in which we obtain self-knowledge motivating the development of Privileged Access.

Boghossian offers an argument of Ryle's to help explain why it is necessary that we don't know our mental states by observing and inferring from behavior. Ryle promoted the idea that the apparent asymmetry described above is just an illusion and we come to know our own mental states the same way we come to know the mental states of others. Boghossian does not offer a complete description of Ryle's theory since he probably figures that most people would find it as absurd as he does but I assume that one of Ryle's claims for the reason the illusion exists is that we know ourselves much better than the way we know anyone else. However, whatever reason Ryle has for his theory doesn't matter because in the actual world when we have an occurring mental state we come to know that state before it has a chance at affecting our behavior in any way. Boghossian offers an example of a thought to show this but I believe that it is better seen with an example of a sensation. Imagine that you are sitting down watching TV and suddenly there is a sharp pain in your foot. Neuroscience has shown us that what happens physically is that at the point of origin of the pain electrical signals are first sent to the brain through the nervous system and then the brain registers a feeling of pain at the point of origin. Since the brain also controls the body's behavior it would be illogical for the brain to send out signals to the body to behave as if it is in pain before the brain even registers that the body has had an experience that causes the pain feeling. If we substitute 'thought' with 'feeling of pain' then this Boghossian quote sums up the moral of the example; "Your knowledge of that occurrent thought could not have been inferred from any premises about your behavior because that thought could not yet have come to have any traction on your behavior." (Boghossian 152) It is a given fact that when a heavy box falls on my toe I feel the pain before I start jumping up and down holding my foot while screaming "Ow! Ow! Ow!" Moreover there is the famous Super-Spartan thought experiment where a whole society is taught to suppress all pain behavior no matter how much



pain they feel. Regardless of the extremeness of the thought experiment it shows that it is indeed possible to feel something and not behave as if you do.

Boghossian moves on to consider other potential forms of inference that could be used to acquire self-knowledge. He uses a tenet of Internalism to help show that any use of inference in obtaining self-knowledge necessarily runs into a problem familiar to many philosophers.

According to Boghossian, the intuition motivating Internalism is that regardless of what is going on in the external world, a person cannot be considered justified in having a belief if from their particular subjective perspective it is epistemologically irrational or irresponsible to hold that belief true. (Boghossian 153) As usual there are thought experiments supporting this intuition just as there are also thought experiments aiming to disprove it. However, one experiment that Boghossian discusses highlights the fact that under Internalism the only evidence available to a person to justify knowledge of their mental state is what is already in their mind because they are only allowed to make a judgment about the mental state from their subjective perspective. Since they are restricted from using anything in the objective world, the person is forced to use other mental states as evidence for justification of the knowledge. This is another way of saying they inferred the knowledge of an occurring mental state from other mental states, past or present.

There is an inevitable problem concerning justification in the field of Epistemology when something is justified by another thing which is not already intrinsically justified itself. Evidently this problem is that if anything is to be used as evidence for justification it must already be justified independent of the thing for which it was intended to provide justification. If a belief in the proposition that  $p$  is justified by a belief in the proposition that  $q$  then the belief that  $q$  must be justified by something else if it is to work as proper evidence for the justification of the belief that  $p$ . An infinite regress is inevitable as long as the path of justification is not brought to a halt

by something which does not require justification by something else. Here is where Internalism is no longer required for Boghossian's argument because he just needed to paint the picture of what it would look like if we were inferring knowledge of our mental states from other mental states. He takes this argument as showing that it is necessary under Internalism that self-knowledge be non-inferential.

Boghossian points out another similar but distinct problem when we consider knowledge of our own mental states instead of only considering knowledge of a proposition. Here Boghossian stipulates that in this scenario to be justified in believing a proposition, the proposition must be justified *relative* to another belief, or other mental state, rather than rest on another belief that carries its own justification. (Boghossian 154-155) Let's say I have a belief that  $r$  and I believe that I have the belief that  $r$  instead of knowing that I believe that  $r$ . For this belief that I have a belief to be justified it must be the case that there is another mental state related to the belief that  $r$  which justifies the belief that  $r$ . This mental state happens to be a belief that  $s$  where the proposition that  $s$  justifies the proposition that I believe that  $r$ . In this scenario, my belief that  $r$  is a justified belief due to the truth of the proposition that  $s$  but we can only be justified in believing that we hold the belief that  $r$  if we also believe that  $s$ . Furthermore we can only obtain justified knowledge that I believe that I believe that  $r$  if we know that we believe that  $s$  and knowledge of the belief that  $s$  justifies the belief that I believe that  $r$ . What this amounts to is that if all self-knowledge were inferential, in order to obtain justified knowledge of a mental state we must already have knowledge of another mental state that justifies the knowledge of the former state regardless of whether the latter state is justified itself. The knowledge of the latter mental state can only be acquired in the same manner which means that it must have been obtained from knowledge of yet another mental state. Instead of an infinite regress of

justification we have an infinite regress of knowledge of particular mental states that must be known before we can have justified knowledge of an occurring mental state. If all self-knowledge is inferential, it cannot be the case that there will be a mental state which does not presuppose knowledge of another mental state because knowledge of the latter state is self-knowledge and must have been inferred from something. At this point Boghossian concludes that the type of knowledge Privileged Access is supposed to yield cannot be obtained through inference since it will inevitably lead to an infinite regress in one way or another.

There is not much to be said in response to Boghossian's first section because whether or not you are inferring from only mental states or from things in the environment as well it is obvious that inference leads to many problems. Boghossian could have even shortened or completely cut out his first section because the Principle of Privileged Access already implicitly restricts inference from yielding inferential knowledge. When Privileged Access is put in terms of direct knowledge, knowledge from inference does not exhibit the property of directness that is required. Any use of inference between the instantiation of a mental state and the acquisition of the knowledge of the state is a mediating factor that must exist between the two stages if the transition is to be possible. For knowledge to be considered direct we must be able to go from the instantiation stage to the acquisition stage without there being any other necessary stages in between. If a particular inference is necessary for the transition then the resulting knowledge must be indirect knowledge. When Privileged Access is put in terms of a priori knowledge, inference can be used if it only involves other things known a priori but the problem of infinite regress is once again apparent.

Boghossian decides to move onto his next section where he attempts to show that the knowledge characterized by Privileged Access cannot be the result of inner observation. He

believes that the true problem here comes from a widely accepted relationist conception of content which dictates that, “the content properties of our mental states and events are determined by, or supervenient on, their *relational* properties.” (Boghossian 157) His first argument against inner observation is the usual claim that in order to say I know my thought involves my *water* concept I need to first be able to know that it does not involve my *twater* concept. Since we can only figure this out by investigating the environment to see which substance I was in the presence of when my thought was instantiated, it is safe to say that I cannot know a priori that my thought was about water rather than twater.

Boghossian concedes that this argument was too swift to account for the whole issue and introduces the relevant alternatives hypothesis. This hypothesis is used with thought experiments whose real-world applications have been outlandishly stretched to include logical possibilities that, as far as we know, have never and will never be actualized. What it states is that when we are using a thought experiment to make a point of argument and this thought experiment asks us to consider a situation where something is the case but it could have been otherwise, the only potential alternate cases that should be offered for consideration are those that are realistic alternatives in the actual world. For instance, twater is not a realistic alternative in the actual world since the existence of a liquid with all the same properties as water except for chemical makeup seems highly implausible for many reasons. The key reason for its implausibility is that the properties of any substance ultimately boil down to its chemical makeup so the different chemical makeup of twater should exhibit some difference in properties. However, Boghossian does not consider this reason for twater’s implausibility and claims that the relevant alternatives hypothesis can be sidestepped rather easily by stipulating situations to be used in thought experiments where the alternatives are relevant.

Let's disregard the inherent problems with twater that question how relevant it can be and imagine Putnam's Twin Earth experiment. The variation Boghossian wishes to introduce is that instead of there being a person on Earth with a twin on Twin Earth, one person is shuttled back and forth between the two planets without their knowledge. If the person stays long enough on a planet then they will acquire any concepts of the planet they did not previously acquire from the other, as long as externalist demands of the acquisition of concepts are met. Boghossian tells us here that this concept acquisition can occur in one of two ways; either the Earthian concepts displace the Twin Earthian concepts upon acquisition and vice versa or upon acquisition they have both planet's respective concepts at their disposal. Boghossian asserts that the latter is the more interesting way to tell the story but since the former is how it's usually described that is what he will use. I agree with Boghossian that the latter is the more interesting, not to mention more realistic, way to tell the story. It is hard to imagine what it would be like to lose a concept we have already acquired just because another similar, but distinct, concept is added to our repertoire. Though, his main argument does not hinge on his choice of the two options so it does not make much of a difference.

Let's consider a person we'll call S who has been shuttled back and forth between Earth and Twin Earth without their knowledge and each time they are on a planet long enough their 'water' concept is displaced with whatever concept corresponds to the planet they are currently on. When S arrives on Earth after such a displacement occurs, S is asked whether they recently had a thought involving a concept very similar to but distinct from their current water concept. Since S is unaware of the existence of twater, Twin Earth, or that the content of his thoughts are dependent on his environment, S most likely would answer 'no'. However, S did in fact have such a thought and his lack of knowledge isn't explained by a lack of memory. Part of the

content of his thought that makes it a thought about twater rather than water was unavailable to S at the time of the thought and when the question was asked. The fact is that the aspect of the content of the thought that makes it a twater thought rather than a water thought is only knowable through empirical investigation. In order to answer the question correctly, S not only needs to know whether they were on Twin Earth or Earth but also have knowledge of chemistry because the original Twin Earth experiment stipulated that this is occurring before anyone can discover the chemical makeup that distinguishes water from twater. (Boghossian 159-160)

Through this Twin Earth variation Boghossian has made twater a relevant alternative, in his eyes at least, and showed that inner observation alone cannot allow us to differentiate one concept utilized by a thought from counterfeits.

However, there is a more serious concern for proponents of Privileged Access than inner observation alone lacking the ability to distinguish a concept from potential counterfeits. This is the argument that there is really no such thing as intrinsic properties when it comes to mental states or events and their content. The current relationist conception of content that he mentioned in the introduction of the paper has it that even non-external factors that determine the content of a mental state or event are relational properties of the state or event. (Boghossian 161) We could also consider the functionalist idea that the only narrow properties that can determine the content of a mental event are the events causal properties. In this regard the content of a mental event would be determined by how the event interacts with other mental events as well as its role in mental functions such as reasoning and contemplation. To illustrate this Boghossian uses a very brief thought experiment comparing thoughts about water to thoughts about gin. In order to know my thought is about water as opposed to being about gin I must know my thoughts causal role is that of R instead of R\* which would be the causal role of a thought about gin. Though

Boghossian does not think that it is possible to acquire this knowledge directly and his reasoning for this revolves around Hume's observation that it is not possible to gain knowledge of anything's causal properties without the use of inference. (Boghossian 161-162)

Boghossian considers an objection to his argument which claims that he has been appealing to a false principle. This false principle consists of the idea that, "in order to know a mental event one must know how things stand with respect to the conditions that individuate that event." (Boghossian 161) Boghossian completely agrees with the objection because his argument would be at fault if it was assuming such a principle but he maintains that his argument relies on a much different assumption. This assumption is that an object's relational or extrinsic properties are not the kind of things that can be ascertained through inspection alone. Boghossian also provides two claims which are supposed to justify this assumption. These claims are that knowing an object's intrinsic properties alone are not enough to know that it has some particular relational property and that the simple inspection of an object can give you at most knowledge of its intrinsic properties. (Boghossian 162) In order to strengthen his argument Boghossian considers another objection that there could perhaps be exceptions to his assumption where inspection of an object can yield knowledge of an extrinsic property. One such exception could be the monetary value of American coins since we seem to be able to know the value of a coin through inspection alone. However, Boghossian believes this is a misconception. The reason we seem to know an extrinsic property of a coin through inspection alone is because some of the coin's intrinsic properties provide this knowledge. To show this Boghossian designs a thought experiment where we are asked to imagine that all coins appear exactly the same and the only thing that dictates what value they have is what kind of mint they were coined at. There are one cent mints, five cent mints, ten cent mints, etc. In this case the only way to know a coin's value is

to know the history of that particular coin which cannot be acquired from inspection alone.

Though Boghossian takes it a step further to argue that even without this thought experiment, the way we ascertain the value of coins appears to be through inspection but it is really inference from the data we acquired from the inspection. (Boghossian 163) We infer from the fact that the coin says ‘ten cents’ on it that the coin is a dime and thus worth ten cents within an American monetary system. Here Boghossian concludes his discussion of why we cannot know the content of our thoughts through inner observation.

The last section of Boghossian’s paper is dedicated to showing how the knowledge of the content of our mental states cannot be based on nothing. Since inference and inner observation are not available as options anymore the last possibility that Boghossian can think of is that such knowledge is the kind of thing where there is no available evidence for its justification.

Boghossian states that when knowledge is gained through observation or through an inference based on an observation it is to be considered a cognitive achievement with a substantial epistemology. There are three examples of knowledge that he considers which do not fall under that classification. Judgments such as ‘I am here now’ are true and justified the moment they are thought without the need of additional evidence. The second example deals with the Kantian thesis that experiencing the world as containing substances is a precondition for experiencing the world at all. In this case the knowledge that the world contains substances is cognitively insubstantial since all that is required for it is that the individual experiences the world. The third example is that of self-verifying judgments like Burge’s Basic Self Knowledge that we discussed in the previous chapter. (Boghossian 165-166) If it is possible to view knowledge of the content of our mental states along these lines then this view would allow us such knowledge. However, Boghossian does not believe that we can view knowledge of mental states in such a way.



There are three characteristics that Boghossian states that knowledge which is not a cognitive achievement would be expected to have which he believes is absent in self-knowledge. The first is that paying more attention to our thoughts should not result in more or strengthened knowledge of the thought than with less attention. For instance, pretty much everything that can be known about the judgment ‘I am here now’ is available immediately and any further attention paid to us does not provide any additional knowledge. The second characteristic is that adults seem to be better at reporting their mental states than children but knowledge based on nothing should be equally available to everyone no matter how well trained their mental capacities are. As Boghossian states, “How is this to be explained if self-knowledge is not to be thought of as an information-sensitive capacity that may be subject to cultivation or neglect?” (Boghossian 167) However, the most important characteristic is that knowledge based on nothing should not be prone to error since it must be true if it is to be considered knowledge in the first place and the fact that it is based on nothing means that no justification is available because it wasn’t supposed to be required at all. Anything prone to error requires justification and anything that does not require justification should not be prone to error. Knowledge based on nothing should also be complete since the only way to complete something that is currently incomplete is to acquire the rest from observation or inference but those options are not available for knowledge based on nothing. As any human being knows, self-knowledge is far from being infallible and we often do not have complete knowledge of our mental states immediately upon their occurrence. If that weren’t the case then what need would the human race have for psychologists and psychiatrists?

While this accounts for most of our mental states of which we want to say we can have direct knowledge, it does not account for Burge’s Basic Self Knowledge which are self-verifying mental states. We saw in the last chapter that judgments like ‘I judge: I believe that thinking

requires concentration' are self-verifying in that having the second-order judgment about the occurring first-order belief provide enough justification in themselves for their truth. The only way the judgment could be wrong is if I did not have the occurring belief that writing requires concentration but since that belief is the object of the second-order judgment then if I did not have the occurring belief I would also not be making the same judgment since the judgment would not have the same object. The first problem Boghossian has with Basic Self Knowledge is that it does not account for our *standing* mental states. He claims that we do not actually have to have the first-order belief in order to have the second-order judgment and without the first-order belief the judgment is no longer self-verifying. While this seems to be ignorant of how Burge characterizes the necessity of the first-order mental states it does not matter because Boghossian's real objection to Burge does not depend on this misconception. Boghossian also believes that Basic Self Knowledge is limited with *occurring* mental states. For the same reason as with standing mental states, we can have the second-order judgment without actually having the occurring first-order propositional attitude. Though Boghossian believes that Basic Self Knowledge was limited with occurring mental states because he believes it is possible for a second-order judgment that occurring simultaneously with a simple first-order thought could be self-verifying. However, Boghossian's real argument against Basic Self Knowledge actually uses compatibilist motivations against a compatibilist argument. Everyday people seem to know their own thoughts without the necessity of generating a second-order mental state about those thoughts. When we think 'writing requires concentration' we do not need to generate any further mental states before we know the content of that thought. (Boghossian 170) So the notion of self-verifying second-order mental states being required for the possibility of us being able to know

our own mental states completely ignores the intuitions of Privileged Access which evoked this whole debate to being with.

Boghossian considers the idea that even though Burge's Basic Self Knowledge does not reach as far as Burge thought, Burge did seem to provide one example where we could know our own thoughts even though its individuation conditions rely on external factors. Though, Boghossian does not believe that Basic Self Knowledge actually constitutes real knowledge. If we consider the Twin Earth slow switching scenario brought up by Burge, when the person is asked whether their thought yesterday was about water or twater it is proposed that they won't know the answer to the question. Boghossian elaborates by stating that it seems in this story that a person will not know tomorrow what they are thinking at this moment but they do know right now what they are thinking at this moment. (Boghossian 171) At this current moment the person is able to have a second-order self-verifying judgment about their thought but no such self-verifying judgment will be available tomorrow. It is worth noting that this statement does depend on Boghossian's claim that Basic Self Knowledge only works with occurring mental states when the second-order state occurs simultaneously with the first. In order for the person to know what they thought yesterday they need to know what their environment was like yesterday. Boghossian thinks there are only two possible explanations for why the person does not remember what they thought yesterday; either they had forgotten or they never knew. However, memory failure is the type of thing that can be removed from thought experiments by careful design so the only option left is that they never knew in the first place.

The problems created by Boghossian are much more serious for compatibilists than those created by McKinsey. Perhaps this is the case because Boghossian claims that when it comes to his beliefs he is a compatibilist so his perspective on the issue is much different than McKinsey's

who is an incompatibilist in his beliefs as well as his arguments. While there are ways around each of Boghossian's arguments, finding those ways without revising our descriptions of content externalism or privileged access requires careful steps and delicate phrasing. The use of inference may not inherently clash with the a priori description of privileged access as long as making the inference does not require any empirical investigation but it does clash when the knowledge of our thoughts we obtain through privileged access is described as direct. Knowledge resulting from an inference had to go through an intermediary step which directly goes against the idea of directness. Therefore any compatibilist theory which hopes to account for the directness feature of privileged access while using inference must explain how directness does not apply to inferential mediating steps.

While most compatibilist theories are developed to accommodate self-knowledge which are not 'based on nothing', some theories do have something to worry about like those which include things like Burge's Basic Self-Knowledge. Such things could still be considered knowledge due to their self-verifying nature but with them we could never know whether the content of that knowledge refers to anything and it could therefore be based on complete fabrications of our mind. It would be hard to call the knowledge that 'I judge that I think the sky is blue' a cognitive achievement, or even true knowledge for that matter, if our concepts of 'sky' and 'blue' are descriptivist concepts that we made up which do not refer to any existing things in the external environment. Compatibilist theories describing self-knowledge which are 'based on nothing' must be able to describe how such knowledge can truly be considered 'knowledge' while also showing how this does not diminish our intuitions about privileged access.

The biggest problem for compatibilists posed by Boghossian's arguments is that it appears like we cannot obtain our knowledge of our thoughts using inner observation. So far no

one has been able to describe a situation where observation alone without the use of empirical investigation has resulted in knowledge of anything other than intrinsic properties. Extrinsic or relational properties are the types of things that can only come to be known through observing how the object or event interacts with other objects and/or events. While we may be able to observe without empirical investigation how a given mental state and the concepts used within it interact with our other mental states and concepts, empirical investigation is necessary to observe how the mental state and concepts used within it interact with objects and events in the external environment. If Boghossian is right that the content of a mental state is determined by its relational properties then we can only assume that part of that content is determined by the relational properties that the mental state bears to external objects and/or events. The only options for compatibilist responses are to argue that knowledge of the content of our mental states does not require knowledge of its external relational properties, that we can have partial knowledge of the content of our mental states through the mental state's relational properties that we can observe without empirical investigation, or to deny either of Boghossian's claims that observation can at most grant us knowledge of intrinsic properties or that the content of a mental state is determined by its relational properties. While there are still outs left open for compatibilists, all four of those options are very difficult to argue.

## Chapter VI

While Burge used the concept of second-order states to describe a subset of self-knowledge to which we can have privileged access regardless of externalism, Kevin Falvey and Joseph Owens used the concept in a different way which accounts for an even larger subset of self-knowledge in their paper, “Externalism, Self-Knowledge, and Skepticism”. Despite the fact that externalism is a metaphysical theory, Falvey and Owens acknowledge that it does appear to have an effect on knowledge of mental states. Although they never use the term ‘privileged access’, you can tell they are using the same principle as we are when they state that externalism seems to undermine the idea that we have *direct* and *authoritative* knowledge of our mental states. (Falvey and Owens 107) The problem for them is that we should have to know our environment has certain features to know the content of our mental states if that content is even partially determined by those features. While Falvey and Owens do not adopt the Cartesian notions of self-knowledge because they recognize the obvious flaws in the Cartesian theories of self-knowledge, they claim that it is not easy to ignore the intuition that when we have an occurring mental state the knowledge we acquire of that state seems to be both direct and authoritative. In between the occurrence of the state and the acquisition of knowledge of that state there is no empirical investigation performed and it seems that our judgments about the states cannot be wrong except in the special cases of self-deception. To quote a phrase that completely exhibits the spirit and motivation behind all Compatibilist theories, “Any philosophy of mind that cannot accommodate and explain these intuitions must be mistaken.” (Falvey and Owens 108)

This particular paper by Falvey and Owens is separated into two parts. The first section deals with how externalism restricts some kinds of knowledge from privileged access and distinguishes between the kinds of knowledge which are and are not restricted by it. In the second section, Falvey and Owens take the set of self-knowledge which they believe they have proven to be applicable to privileged access, as well as the theory behind why it is not restricted by it, and use it as an argument against skepticism. We are only going to discuss the first section here because while defeating skepticism would be tremendous for the Compatibilist cause since the idea of having a priori knowledge of the external environment would no longer be considered an absurd conclusion; it is way too big of an issue to be taken on as a side project. Skepticism has plagued every area of philosophy since the first claim was ever made so it seems much more productive to push it to the side until it is the last issue that needs to be addressed. Besides, if Falvey and Owens provide an argument which frees more knowledge of mental states from the restriction of privileged access than Burge's Basic Self Knowledge then we have progress on the Compatibilist front despite skepticism. They in fact do provide such an argument which distinguishes between two kinds of knowledge of the content of mental states that most philosophers lump together. One of these kinds of knowledge is restricted from privileged access for many of the same reasons that Incompatibilists argue all knowledge of mental states are restricted. The other kind, they argue, is not restricted for reasons related to the concept of second-order states and how their content is determined.

Falvey and Owens begin their discussion with the now very familiar Twin Earth thought experiment developed by Putnam. Sparing the reader from another description of the experiment all that is important is that whenever the person on Earth and their twin on Twin Earth express a thought using their respective 'water' concepts, they are expressing thoughts with different

content even though they are physically identical. Here is where Falvey and Owens offer their distinction between two kinds of knowledge often used in similar thought experiments. When we say our knowledge is direct and authoritative there are two different ways this can be taken:

- (1) An individual knows the contents of his occurrent thoughts and beliefs authoritatively and directly (that is, without relying on inferences from observation of his environment). Call this kind of knowledge *introspective knowledge of content*.
- (2) With respect to any two of his thoughts or beliefs, an individual can know authoritatively and directly (that is, without relying on inferences from observation of his environment) whether or not they have the same content. Call this kind of knowledge *introspective knowledge of comparative content*. (Falvey and Owens 109-110)

Falvey and Owens readily concede that introspective knowledge of comparative content is incompatible with externalism but this is mostly due to the fact that they believe such knowledge is impossible regardless of whether externalism is true. However, Falvey and Owens also believe that there is no reason to think that introspective knowledge of content suffers from the same problem as (2) because they think all Incompatibilist arguments provided so far do not make a distinction between (1) and (2). (Falvey and Owens 110)

Falvey and Owens present two different thought experiments to illustrate why introspective knowledge of comparative content is incompatible with externalism. In the first we are asked to imagine someone named Rudolf who is well acquainted with cilantro to the point that he knows it is an herb, it is used in Mexican dishes, and it has a distinct smell. He often has thoughts like, "Cilantro should be used sparingly," which he expresses out loud. Rudolf also uses dried coriander often when he cooks and it is stipulated that he is just as familiar with coriander



as he is with cilantro. Similarly, he often has thoughts like, “Coriander should be used sparingly,” which he expresses out loud. (Falvey and Owens 110) Although, Rudolf is not aware of the fact that cilantro and coriander are two distinct names pertaining to the same concept.

The question is whether Rudolf can know his two thoughts are the same without empirical investigation? Since the herb that both terms refer to is a natural kind and natural kind concepts are the least controversial of all the concepts to which externalism is supposed to apply, it seems like Rudolf is not capable of comparing his concepts of cilantro and coriander without the knowledge that is only available in the environment that they are the same. (Falvey and Owens 110-111) Although they never state it openly, I assume that Falvey and Owens do not think Rudolf can use his introspective knowledge of the qualitative aspects of his mental states using each concept to compare the concepts. The reason is that even if he could compare the taste and smell of each concept to determine that they are so similar to the point of being identical, he cannot know for sure that cilantro and coriander are both actually the same herb without at least looking up the words in the dictionary. For all Rudolf knows, cilantro and coriander could be two distinct herbs which smell and taste almost identically but at least one has different properties which cannot be discovered without the correct tools. For example, one of the two could cause allergic reactions in some that the other doesn’t cause or the flavor of one of them may disperse throughout the food when cooked while the other holds the flavor regardless of how it is cooked. Though, it is clear in the next paragraph that Falvey and Owens acknowledge that there are introspectible clues available that Rudolf could have tried to discover to help him compare the two. This is because they use the next thought experiment to show the same result without the presence of introspectible clues.

The other thought experiment that is meant to show that introspective knowledge of comparative content is incompatible with externalism is a variation of the Twin Earth experiment similar to Boghossian's slow switching version of the experiment where our subject, Susan, is shuttled back and forth between Earth and Twin Earth without her knowledge. Each time a switch is made her concepts are replaced by the concepts of the social population on that planet once she is there long enough. Susan believes that she has never left Earth and her concept of 'water' is the same as all other people on Earth, but unbeknownst to her at some points in her life when she believed she expressed a thought about water it was actually about twater. If we were to ask Susan while she is on Earth, and her Earth concepts have replaced her Twin Earth concepts, when she says the sentence right now 'water is a liquid', is she expressing the same thought as she did a year ago when she said the same exact sentence? Since Susan has never been aware of the fact that she had been on any other planet than Earth she will answer yes. However, if she was in fact on Twin Earth last year when she uttered the sentence she would be wrong because her 'water' concept would have been different. Furthermore, the information that would allow her to answer correctly is only available in the external environment through empirical investigation. Therefore, Falvey and Owens conclude that if externalism is true then (2) must be false.

Though, Falvey and Owens do not believe this presents a new problem for externalism because they think that there is reason to believe that (2) would be false whether or not externalism is true. The reason for this rests on yet another thought experiment. We are asked to consider two sentences:

- (i) Nobody doubts that whoever believes that Mary is a physician believes that Mary is a physician.

- (ii) Nobody doubts that whoever believes that Mary is a physician believes that Mary is a doctor.

We are then presented with philosophers Benson Mates and Alonzo Church who have differing views of these two sentences. When Benson Mates utters both sentences he believes he is expressing different thoughts but when Alonzo Church utters both sentences he believes that he is expressing the same thought. Each philosopher understands that each sentence expresses a thought through the English language and they are both aware of the meanings of all the terms involved within the English language. One of them must be wrong but according to Falvey and Owens, the mistake made on either side is not likely to be a result of introspective failure.

(Falvey and Owens 113) The true message of this thought experiment revolves around our ability to make a judgment about which philosopher is right in this disagreement. We are in the same position as both Mates and Church in that we fully understand the content of both (i) and (ii) as well as the meanings of all the terms involved in the English language. We believe that the terms 'physician' and 'doctor' are synonyms so both (i) and (ii) should express the same thing.

However, through introspection alone it is extremely unlikely that we can find enough evidence to make any kind of judgment on the matter. Even if we believe that the thoughts expressed by both (i) and (ii) have the same content, how do we know we are not mistaken in this belief if we don't do any empirical investigation into the practice of the English language and how it is commonly used today by those who practice it? For all we know at some point in time since we learned that 'doctor' and 'physician' are synonyms, someone could have developed an argument as to why the terms shouldn't be synonymous and successfully changed how they are defined in the English language. "Even if one knows what one is thinking at a given time, and knows what

one is thinking at a later time, it may be necessary to know something about one's environment in order to know whether these two thought contents are identical." (Falvey and Owens 113)

While I agree with Falvey and Owens that due to all the previous thought experiments introspective knowledge of comparative content does seem to be incompatible with externalism, I find it hard to see how the last thought experiment is supposed to show how (2) is false independent of externalism when externalism was clearly not taken out of the picture. The problem was only generated by the idea that Mates and Church both believe that the thought expressed by the sentences is the same thought expressed by the sentences in English. They cannot both be right because the definition of a physician in English cannot both include and not include doctors. This means the information which dictates who was wrong can only be found through assessing the current state of the English language and how each term is defined within it. Nowhere in this paper did I read that Falvey and Owens were restricting the social environment from the kinds of external environment to which externalism applies and the linguistic practices of those who speak a common language definitely falls under one's social environment. Furthermore, they state that, "while much of this information is logico-philosophical in character, it is not plausible that it can be acquired independently of a serious empirical investigation into linguistic practice." (Falvey and Owens 113)

So far the idea of empirical investigation has been the quintessential process which describes the hold that externalism has on the concepts used in our mental states. The reason we are not supposed to be able to have direct, authoritative knowledge of our own thoughts, according to Incompatibilists, is because externalism entails that some aspects of our mental states were determined by the environment and therefore only knowable after sufficient empirical investigation. If, in this thought experiment, the relevant knowledge which would

answer whether it was Mates or Church who was wrong can only be found through some empirical investigation it is extremely implausible that externalism does not play any role in the experiment. Yes this is a different perspective on externalism than we have commonly been looking at because it is not the physical environment determining the content of the thoughts expressed by the sentences in English but rather the collaboration of humans over many years. However, the external environment for any individual also includes the other living beings in their environment because to that individual the only difference between the living beings and inanimate objects in the environment is that the living beings have a conscious mind directing their behavior. Therefore, the only difference between the social environment and the rest of the external environment is the fact that the social environment is molded by conscious minds. Just as natural kind objects in the environment determine the content of their respective natural kind concepts, whatever the content is of a thought expressed by a sentence will have been determined by how the terms involved have been commonly defined by those who speak the English language. So I conclude that unless Falvey and Owens have a good argument for why externalism does not apply to the social environment then this last thought experiment does not prove that (2) is false independent of externalism.

The problem discussed in the last two paragraphs is significant because if I am correct, it is possible that introspective knowledge of comparative content isn't false independent of externalism which turns the tables by pointing to externalism as the plausible location of the problem instead. However, when it comes to the compatibility of externalism and privileged access the entire discussion of introspective knowledge of comparative content is irrelevant because if we can show the introspective knowledge of content is compatible with externalism then the desired progress we wanted for compatibility theories is reached. Falvey and Owens

consider two separate arguments for why introspective knowledge of content is compatible with externalism. Unfortunately the first argument they use as a response to content-skeptics utilizes their conclusion that we necessarily lack introspective knowledge of comparative content in their reasoning so anyone who agrees with me that they did not actually prove this will find the argument unsatisfactory. Lucky for us they also reject that first argument as being insufficient since the manner in which the content-skeptic phrases the argument can make the lack of introspective knowledge of comparative content irrelevant to the situation at hand. They then follow up their rejection with a response that is undeniable when upholding their premises.

In order to begin their discussion on why externalism does not entail a rejection of introspective knowledge of content, Falvey and Owens decide to start from scratch. Is externalism incompatible with the idea that we can know our own mental states directly and authoritatively when we are assuming that we lack the ability to compare the content of our mental states through introspection? Given the existence of the issue at hand it would appear that the answer to this question is yes. (Falvey and Owens 113) This answer is often due to the reasoning of the conclusion of thought experiments similar to Putnam's Twin Earth. Essentially, they all boil down to the idea that if someone were to be said to have knowledge of the content of one of their occurring mental states using a certain concept then they must be able to know that the concept utilized is that particular concept rather than some similar alternative concept. Since it is very simple to design thought experiments where the information needed to distinguish between the potential concepts is necessarily unavailable, introspection alone cannot accomplish this feat and therefore we do not have introspective knowledge of content which is both direct and authoritative.

The usual compatibilist response to this argument by those who have not yet read Paul Boghossian is to present the Relevant Alternatives argument. Falvey and Owens present this argument for two reasons: they are starting from scratch so it is useful for the reader to see the path that the debate has taken over the years but more importantly a form of this argument becomes a crucial point in their reasoning for why we can have introspective knowledge of content with externalism. The Relevant Alternatives argument proceeds by claiming that even though one must be able to rule out certain alternative hypotheses in order to have knowledge, it is not the case that one must rule out every alternative hypothesis that is logically possible. (Falvey and Owens 114) It is when the conditions that determine what counts as knowledge are so strict that they require that everything logically possible must be accounted for that it is possible for Socrates' paradox, "The only thing I know is that I know nothing", to actually carry some real truth. For any given situation the set of logically possible alternative hypotheses is virtually infinite because it is only limited by what the mind can imagine which doesn't necessarily contradict itself. This set includes alternative situations where the universe is governed by different physical laws or even concepts of space and time that are different from whatever the nature of space and time are in our actual universe. If the conditions for knowledge really were as strict as this then the incompatibilist making the argument would be right that we don't have knowledge but their judgment could not constitute knowledge either since they also cannot rule out every logically possible alternative. You may say that I don't know that my occurring thought is about water because I can't rule out that it wasn't about twater. However, if we are considering all logically possible alternative hypotheses then you can't rule out the possibility that I do know it wasn't about twater because perhaps I am an alien or some superhuman with much more evolved mental capacities who has the equivalent of what people

call 20/20 vision in regards to introspection and also have the ability to have introspective knowledge of comparative content when no other human can. The simple fact is that when we do not restrict the conditions for knowledge to only account for relevant alternatives everybody loses because knowledge is virtually impossible for those whose mental capacities are limited by the current stage in the evolution of the human mind.

Falvey and Owens make this point through first postulating a novel thought experiment and comparing it to Twin Earth. Suppose we imagine a man name Tom who is driving through rural country roads where there are many barns to be spotted along the way. Tom has great eyesight, knows the correct use of the word 'barn', and the weather conditions are perfect for making visual judgments. So when Tom sees a barn and judges it to be a barn one would be hard-pressed to claim that Tom does not know that the object he is viewing is a barn. However, what if we were to learn that the area Tom is driving through is full of structures which are not barns but were made to look like barns and they fulfill their purpose extremely well? In this case we would actually want to claim that Tom does not know the object he is looking at is a barn because the possibility that it is a barn facsimile would undermine the knowledge. If Tom were an expert in detecting barn facsimiles that would be another story but since Tom is an average person and barn facsimiles are meant to trick the average person into thinking it is a barn it is likely that Tom cannot rule out the alternative that it is a facsimile.

This thought experiment seems counterintuitive because their immediate goal is to show that irrelevant alternatives should be disregarded instead of showing an instance where an alternative that previously didn't enter into our minds becomes relevant. The reason is that they wanted to highlight the difference between what is and isn't relevant. The idea of twater that comes with Twin Earth has much less relevance as an alternative to water than the idea of barn



facsimiles because the postulation of someone wanting to construct an object with the purpose of tricking people into thinking it is something else is much easier to digest than the postulation of an exact duplicate of Earth where the only difference is the chemical makeup of the liquid labeled ‘water’. According to Falvey and Owens, we should require that Susan be able to distinguish water from turpentine and gin but twater should not be necessary since in the actual world it is nothing but fiction. (Falvey and Owens 115)

While the original Relevant Alternatives argument worked for awhile because it kept the infinite realm of logical possibility out of the discussion, the flexibility in the design of thought experiments brought the problem back with a vengeance. Since the creator of a thought experiment specifics all the premises, assumptions, and conditions of the experiment it is not very hard to design an experiment where certain logical possibilities can become relevant when they previously were not. If we considered the version of Twin Earth developed by Paul Boghossian where Susan is shuttled back and forth between planets without her knowledge then twater does become relevant because by stipulation it is an actual liquid she encounters from time to time instead of being just fiction. According to Falvey and Owens, Boghossian would claim that Susan doesn’t know whether her thought is about water or twater once twater is made into a relevant alternative. To illustrate they provide a quote of Boghossian that he makes when he discusses a thought experiment similar to the switching version of Twin Earth above where we are considering two possible realities that take place on Earth where the only difference is how the condition *arthritis* is defined. He calls the concept with the incorrect definition *tharthritis*. Boghossian states that,

“[the subject] has to be able to exclude the possibility that his thought involved the concept *arthritis* rather than the concept *tharthritis*, before he can be said to know what

his thought is. But this means that he had to *reason* his way to a conclusion about his thought; and reason to it, moreover, from evidence about his external environment which, by assumption, he does not possess. How, then, can he know his thought at all? – much less know it directly?” (Falvey and Owens 115)

These conditions for knowledge are hard to overcome, especially in the face of any logical possibility being relevant. Although as long as it can be stipulated in a thought experiment in the correct way, Falvey and Owens are not very concerned because the nature of second-order mental states provides a route of argument that helps sidestep the inclusion of implausible alternatives which are only relevant through specific stipulations in a thought experiment.

Falvey and Owens start by stating that, “*q* is a relevant alternative to *p* if *q* is incompatible with *p*, and the possibility that *q* obtains is relevant in the context.” (Falvey and Owens 116) According to them, while Boghossian doesn’t make this claim directly they believe he is holding onto a Relevant Alternatives principle like the one that follows:

- (RA) If (i) *q* is a relevant alternative to *p*, and
- (ii) S’s belief that *p* is based on evidence that is compatible with its being the case that *q*, then
- S does not know that *p*.

Falvey and Owens believe that this not only appears to be a good principle in which to base our conditions for knowledge but it seems to be the principle we use on a day to day basis. However, they believe that the motivation behind the principle requires a closer look. Revisiting the thought experiment with Tom, Falvey and Owens think that it is not really the lack of evidence

that the object in front of him is not a barn facsimile which causes us to think that Tom's belief that the object is a barn doesn't constitute knowledge. Instead it is the idea that if Tom was in the counterfactual situation where the object in front of him actually was a barn facsimile, Tom would still believe it was a barn nonetheless. This leads them to the postulation of a different Relevant Alternatives principle which they believe is more in line with our intuitions pertaining to what relevant alternatives must be ruled out before we can have knowledge. It is as follows:

(RA\*) If (i)  $q$  is a relevant alternative to  $p$ , and

(ii) S's justification for his belief that  $p$  is such that, if  $q$  were true, then S would still believe that  $p$ , then

S does not know that  $p$ . (Falvey and Owens, 116)

Although they never state it explicitly, I believe one of the reasons Falvey and Owens think this principle captures the intuitions behind the conditions for knowledge in these types of situations is because it is impossible to tell whether S would have the same evidence when their belief is correct as they would when their belief is not. S's evidence for their belief may be compatible with a relevant alternative but there is no way of knowing if the evidence would have been the same if the alternative situation had obtained. Perhaps in the alternative situation, S would not have had the belief they did in the actual situation because they wouldn't have had the same evidence to justify it so in turn the belief would not have been developed.

This brings us to Falvey and Owens' next point. They claim that the nature of the evidence used to justify beliefs will often be a reliable indicator of how prone their belief is to error. When the evidence for a belief is perceptual it will follow that (ii) in (RA) entails (ii) in (RA\*). (Falvey and Owens 117) Tom's belief that there is a barn in front of him rests solely on

his visual perceptions of the object. Since his visual perceptions of the object would be the same if instead it were a barn facsimile then the evidence would be the same across situations which satisfies (ii) in both principles. Whenever evidence is perceptual, for it to be compatible with another situation obtaining strictly means that in the alternate situation the perception will provide evidence which offers the same degree of justification towards the same conclusion as in the actual situation. Since one would expect at least small differences in the perceptual experience of different objects even when one is crafted to be identical to the other, the same exact perception is not necessary as long as it justifies the same conclusion to the same degree.

Though when we consider cases where the belief is not about our perceptions but rather our own mental states, (ii) in (RA) does not entail (ii) in (RA\*). Falvey and Owens think that this conclusion can even be illustrated using the Twin Earth switching experiment. Nothing in Susan's evidence that her thought was about water can rule out the possibility that she was instead on Twin Earth thinking about twater. According to Falvey and Owens, "such a situation would be evidentially indistinguishable from her actual situation." (Falvey and Owens 117)

However, that does not mean that if Susan were on Twin Earth with a thought about twater that she would believe she was on Earth and her thought was about water. Falvey and Owens claim that Susan would not have the latter belief because, according to externalism, the content of second-order beliefs are determined by the environment the same way first-order beliefs are. Let it be said that this goes for all second-order mental states in general; Falvey and Owens must have just decided to use 'belief' because it is much easier to follow the discussion when you are only mentioning a specific kind of mental state rather than all of them in general. When Susan is on Twin Earth it is not possible for her to have a belief about water because she doesn't have the Earthian water concept when on Twin Earth. This argument depends greatly on their assumption

that in thought experiments where we switch someone back and forth between different environments, concepts that are supposed to correspond to another in the alternate environment but have slight differences will replace each other once the person has remained in the environment long enough. Surprisingly enough, this assumption is not very controversial among philosophers in the field despite the plausible alternative that instead of replacing each other we just acquire a new concept once we stay in the environment long enough which has the same label as a concept we already possessed. We will return to this at the end of the chapter.

Falvey and Owens believe that this is enough to show that someone can hold externalism to be true and endorse (RA\*) without having to accept the argument that we do not have introspective knowledge of content. As long as the environment determines the content of our second-order mental states the same way it does our first-order states, it is not possible for the content of my second-order state to include concepts that differ from the concepts in the first order state. Furthermore, as long as the content of any second-order mental states is derived from the content of the first-order mental state to which it pertains then it is not possible for the second-order state to be wrong. Even when Susan is under the impression she is on Earth when she is actually on Twin Earth, Susan would still be able to know the content of her thoughts according to (RA\*) since (ii) isn't satisfied and thus neither water or twater are relevant alternatives to each other because it is not possible for Susan to be in possession of both concepts at the same time.

While we might hesitate to call this information knowledge since it is based on the assumption that the water and twater concepts replace each other, it certainly would be justified enough to be knowledge in terms of correctness. The content of the second-order state will be the same content of the first-order state and, unless it is a case of self-deception, the second-order

state will use the content to accurately describe the first-order state. Whether or not Susan is on Earth or Twin Earth, her second-order belief 'that she is thinking water is a liquid' will take the content of 'water is a liquid' from her first-order thought. Even though which planet she is on will matter towards what the content of 'water is a liquid' actually is, her lack of knowledge concerning which planet she is on does not make a difference to what the content of the second-order belief will be. The content of the second-order belief will always be the content of the corresponding first-order mental state and the content of the first-order state will always be dictated by the concepts utilized in that state. While the set of available concepts and their possible content is determined by the environment we inhabit, the accuracy of our second-order mental states is left untouched because the only content that can be derived from the first-order state is what is already in the state and nothing else.

Falvey and Owens believe that the fact that the content of second-order mental states is determined by the environment provides an argument against those of Anthony Brueckner's content-skeptics. When I claim that I know some water is dripping from a faucet, the content-skeptic will respond by claiming that I believe water is dripping but I do not know water is dripping because by closure principles knowing that would entail that I know I am not a brain in a vat and essentially everything that would come as a result of the physical reality being real. The closure principle being referred to is the principle that knowledge is closed under known entailment which dictates that if an individual knows that  $p$  and that  $p$  entails  $q$ , then they know that  $q$ . (Falvey and Owens 119) The content-skeptic can proceed to claim that we don't know the contents of our thoughts for similar reasons. If I actually had knowledge that I am thinking water is dripping then by closure I would also have knowledge that I am not thinking twater is dripping. According to Brueckner, if I were on Twin Earth without knowing it then things would

seem exactly as they do now on Earth so I can't know that I am not thinking twater is dripping.  
(Falvey and Owens 119)

This is where Falvey and Owens use their conclusion that we do not have introspective knowledge of comparative content. Without such knowledge it is possible for someone to know that they think water is dripping without in turn also knowing that they are not thinking that twater is dripping. The reason is that in cases where we don't have the relevant knowledge to compare a water thought to a twater thought there is no way for us to have the knowledge that it is a water thought which would result in the knowledge that it is not a twater thought. The closure principle is not satisfied because we lack the knowledge that  $p$  entails  $q$ . (Falvey and Owens 120) However, Falvey and Owens concede that if the content-skeptic were to phrase their argument differently the closure principle would no longer makes a difference.

The skeptic could grant us the belief that we are thinking that water is dripping but then ask us to imagine a possible world where our environment contained a liquid that resembled water in every perceivable way except that it isn't water. I assume that the environment doesn't also contain water because Falvey and Owens add that our thought in such a situation would not be about water due to externalism. Since water and twater concepts are different which entails that thoughts involving each are different then in order to know that a thought is a water thought rather than a twater thought we must at least know that it is not a twater thought.(Falvey and Owens 120) It would seem that with the inclusion of possible worlds the lack of introspective knowledge of comparative content does not save the compatibilist because imagining a possible world allows us to entertain the idea of a concept that isn't in our actual environment. The justification for this reasoning is based in the idea that I cannot know I am not a brain in a vat because if I were a brain in a vat everything would "seem" exactly the same as it does right now.

Specifically I would believe that I was not a brain in a vat when in such a situation I would be.  
(Falvey and Owens 121)

There is a problem with skeptical arguments that take these routes. When it comes to counter-factual situations about our perceptions or our qualitative mental states we can confidently claim that we would have the same perceptions in the counter-factual situation that we did in the actual one. However, when we are discussing the content of our mental states there is no way to make a claim that the content of those states, our beliefs in this particular situation, would be the same in the counter-factual situation as they are in the actual one. This is precisely due to the fact that the content of our second-order mental states are determined by the environment in the same way that our first-order mental states are. (Falvey and Owens 122) If we are asked to imagine what the content of our beliefs would be in a situation where our environment is different, the content of our beliefs would change along with the environment since it is what determines the content of those beliefs. Even if qualitatively everything “seemed” the same in the counterfactual situation there is no way you can maintain externalism and expect the content of some mental states to be consistent while changing the environment. We know the content of our second-order mental states because externalism guarantees that the content of our first-order mental states could not be anything other than what it actually is and that content is directly transferred to our second-order mental states. “Just as I cannot think that water is wet unless my environment satisfies certain features, so I cannot think that I am thinking that water is wet unless my environment satisfies the same features.” (Falvey and Owens 123)

While Falvey and Owens did not allow for us to have introspective knowledge of comparative content for our mental states, they did provide an argument which allows for us to have introspective knowledge of content while maintaining externalism at the same time. This is



a tremendous tool for compatibilists to respond to content-skeptics but in the end all they really did was sidestep the issue. Showing how the traditional skeptical arguments can't actually get off the ground does not truly answer the question of how you can know your water thought without knowing that it isn't a twater thought. Just because there is no way for certain mental states to remain the same in a counterfactual situation as it is in an actual one does not change the fact that to know that a given mental state has a certain identity you must know it does not have a certain alternative identity. Essentially this all boils down to the issue of whether a pair of concepts with the same label would replace each other when environments are changed or if the pair could coexist regardless of the current environment that the individual is inhabiting. When the concepts must replace each other then Falvey and Owens are correct that the traditional skeptical argument can't get off the ground but if the concepts could coexist then is it possible for the content of mental states using one of those concepts to remain consistent with a relevant change in the environment. Susan could have a thought about water on Twin Earth and vice versa. Though, at the same time consistency in the mental state is not a guarantee in such a situation because it could have been the environment that evoked the idea. In such a case the change in environment in the counterfactual situation would change the content of the state. Either way it is solely because of the assumption that the concepts must replace each other that their arguments have any real traction. If it turns out that the concepts do replace each other than Falvey and Owens have done a superb job for the progress of the compatibilist cause but until the alternate option has been ruled out the fact that their response to content-skeptics only sidesteps the issue cannot be ignored. Until Falvey and Owens can actually answer the question of how you can know your thought has one identity without being able to know it does not have some alternative identity, their compatibilist theory must be taken with a grain of salt. To borrow a moral from the

authors themselves, any philosophy of mind that cannot account for this must be mistaken, or at least has a lot of holes in it.

In his book *The World Without, The Mind Within*, Andre Gallois also presents a solution for Boghossian's argument which is similar to Falvey and Owens solution of distinguishing between introspective knowledge of content and introspective knowledge of comparative content. Boghossian believes that he has established two claims:

- (a) An individual in the switching situation does not know which of these beliefs she has:  
the water belief or the twater belief.
- (b) An Individual in the switching situation does not know that she believes water is tasteless. (Gallois 179)

Boghossian spends the majority of his paper justifying (a) and then argues that (b) follows directly from (a). While Gallois does not make any claim as to whether Boghossian actually established (a), he does wish to refute the idea that (b) necessarily follows from (a).

Let's grant Boghossian the claim that it does not seem plausible to think that an individual in the switching situation would know whether their thought was about water or twater. Gallois does not contend with the notion that someone in the switching situation can only know that they believe water is tasteless if they know which belief is expressed when they use the proposition 'Water is tasteless'. However, he argues that even when we grant Boghossian those claims it does not have to follow that (b) is true. To justify this claim Gallois presents a simple thought experiment where we are asked to imagine that we are looking at a table. We are then asked if we know which object is in front of us. According to Gallois, this question is unanswerable unless the relevant contrast is specified. (Gallois 181) By relevant contrast he

means which set of objects to which we are supposed to be comparing the object in front of us. For instance, I know that the object in front of me is a table rather than a chair but I may not know that the object is a table from this room rather than a table from the other room. When we use the idea of relevant contrast in the switching situation, the individual may not know that their belief 'Water is tasteless' is a water belief rather than a twater belief but they do know that their belief expresses the idea that water is tasteless rather than expressing the idea that grass is purple. (Gallois 181)

Gallois believes that the second of these contrasts is what is relevant when it comes to having knowledge of the content of our thoughts. The reason for his belief is that it is obvious that someone in the switching situation who does not know whether they are having the thought that water is tasteless as opposed to the thought that grass is purple cannot be said to know the content of their thought. However, what is not as obvious is the idea that someone having the belief that 'Water is tasteless', but doesn't know whether it is about water or twater, does not know that they believe water is tasteless. Since upon the occurrence of our thought 'water is tasteless' we seem to be able know that it is that thought opposed to the thought 'grass is purple' then we appear to be able know the content of our thoughts. Our lack of the ability to know that we are not in a Twin Earth switching situation may not allow us to know whether our thought uses a water concept instead of a twater concept but this should not take away from the fact that there are a innumerable amount of concepts that we do know are not present in the thought.

We will look at a different argument of Gallois's which pertains to McKinsey's argument in Chapter VIII. The reason his argument against Boghossian is included in this chapter is because it is very similar to Falvey and Owens's argument so the best way to notice the slight

differences is to put them next to each other. Both arguments draw on the idea that just because we cannot know that our thought uses a given concept rather than some relevant alternative concept it doesn't mean that we cannot know the content of the thought. Falvey and Owens argued that we cannot have a priori knowledge that our thought uses a particular concept rather than some other concept but we can have a priori knowledge of the content of our thought. Gallois, on the other hand, argues that we can have a priori knowledge that our thought uses a particular concept rather than the set of irrelevant alternative concepts. While the only difference seems to be that Gallois allows for some introspective knowledge of comparative content where Falvey and Owens do not allow for any, the implications of this small difference are much bigger.

Gallois's argument is prone to some objections that Falvey and Owens' argument is not because it must account for how we can differentiate between concepts that are not relevant alternatives to each other but we cannot differentiate between concepts that are relevant alternatives to each other. The only instances where I can imagine that such a claim would make sense are those where the information that distinguishes a given concept from a relevant alternative is restricted from the individual in question. The Twin Earth thought experiment is one such instance. Actual examples of this occurring are very rare if there even have been any. Such an example would consist of two distinct objects or events where all observable properties are the same and what differentiates them is information that is either completely inaccessible or only inaccessible at a particular time period where we lack the technological or biological ability to make the necessary observation. Since most substances have properties that we have been able to observe throughout our history then it is hard to imagine such cases. However, one could imagine that prior to developments in metalworking many people could not observe differences

between some types of metals. In such a case a person may not be able to differentiate a thought about aluminum from a thought about nickel because at that point in time they are relevant alternatives to each other but the person may still be able to differentiate their thought about aluminum from a thought about oxygen.

In contrast, Gallois's argument does a better job at addressing how it is possible for us to have a priori knowledge of the content of our thoughts since we can at least partially know the identity of our concepts used in our thoughts. In Falvey and Owens' argument we could not know that our thought 'water is wet' is not the thought 'grass is purple' which seems rather ridiculous. In reality when we have the thought 'water is wet' we may not be able to completely rule out that we are not in a Twin Earth switching situation so we cannot know that the thought is not actually about twater but we do seem to be able to know that the content of the thought is completely different from the content of the thought 'grass is purple'.

Gallois's argument also does not require that a water and twater concept must replace each other in the Twin Earth switching situation because in Gallois's argument our ability to know the content of our thoughts does not revolve around the idea that the content of our thoughts would change with the appropriate change in our environment. It is consistent with Gallois's argument that we could have both a water and a twater concept at the same time no matter which environment we are currently inhabiting. Furthermore, there is reason to believe that the concepts should not replace each other. Let's imagine a development on the Twin Earth switching scenario where on both Earth and Twin Earth humans have developed the ability to determine the chemical makeup of the substances they both refer to as 'water'. Now when Susan is transported back and forth between both worlds she learns that each water substance has a different chemical makeup. In such a case it seems unlikely that Susan would lose her water

concept on Twin Earth due to it being replaced by a twater concept. What seems more likely would be that Susan would keep both concepts when she learns that they are actually distinct substances. Woolly Mammoths no longer exist in our environment and yet we still have a concept that refers to them. So why would both concepts replace each other when Susan is unaware of each substance's chemical makeup? Prior to this development it seemed like externalism determined that both concepts replace each other in different environments since Susan uses the same label for both and in a twater environment her ability to have a thought using a concept with the label 'water' has a necessary relation to the twater that exists in her environment. After this development it appears like Susan's own mind is what determines her concepts are sorted and labeled since the only change in her environment that caused the distinction between her water and twater concepts had nothing to do with the referents of the concepts but instead with our observational abilities. A change in our observational abilities is not enough to go from it being necessarily the case that two concepts replace each other to their potential coexistence. Rather it is more likely that both water and twater were the referents of Susan's water concept and once the discovery of their difference was made the one concept was split into two.

The fact that Gallois's argument against Boghossian is consistent with the coexistence of water and twater concepts no matter which environment an individual currently inhabits makes the argument much more resilient than the argument of Falvey and Owens. However, we will see in Chapter VIII that this argument of Gallois's is part of an overall theory about the compatibility of content externalism and privileged access which has flaws despite its strength against objections. The main flaw being that in the theories claims that we cannot know the epistemic status of our concepts as a natural kind or a descriptivist concept it diminishes parts of privileged access which cannot be diminished without changing the principle altogether. If we recall from

the first chapter, true compatibilist solutions must keep both content externalism and privileged access consistent with the intuitions that led to their development in the first place.

## Chapter VII

In her paper “Externalism and A Priori Knowledge of the World: Why Privileged Access is Not the Issue”, Maria Lasonen-Aarnio attempts to use the Incompatibilist’s own argument against them by showing that the assumptions that must be made in order to validate such arguments entail that the premise which corresponds to the principle of Privileged Access plays an unnecessary role in the argument itself. If she is correct then the ‘Incompatibilist’ argument is no longer an incompatibility argument but instead just an argument against content externalism. She divides her paper into three sections: in the first she provides and explains all the necessary factors of her argument. In the second section she shows how the current Incompatibilist arguments require assumptions that back their advocates into dangerous corners where the only ways out are to appeal to controversial claims. In the third section, Aarnio explains how these controversial claims take the Incompatibilists down paths that lead to consequences that change the classification of their position altogether. In this chapter I will provide an overview of her central argument while explaining the necessary factors of her argument. Afterwards we can see how the argument actually plays out and decide how well it actually fulfills the purpose for which it was designed.

Aarnio’s central argument is somewhat simple but she makes very swift moves around complicated issues while elaborating on it. Therefore, to make everything easier to follow I will provide an overview of her central argument before diving into how the argument plays out. Aarnio starts her first section by explaining Michael McKinsey’s argument from his paper that we discussed in the second chapter. To reiterate, McKinsey argues that content externalism is



incompatible with privileged access due to the absurd conclusion present in the following argument:

- (1) Suzy can know a priori that she is thinking that *p*.
- (2) The proposition that Suzy thinking that *p* logically implies the proposition that *E*.
- Therefore,
- (3) Suzy can know a priori that *E*.

The use of logical implication in (2) is a controversial issue that McKinsey believes he successfully argued for its truth by showing that metaphysical entailment leads to a conception of wide states that Burge, being one of the developers of the terms wide and narrow mental states, would not agree with. Aarnio also states that sometimes the argument above is put into terms where externalism does not have to be a logical truth but must still be knowable a priori. The only difference between the two arguments lies in premise (2) which is instead substituted with the following premise:

- (2\*) Suzy can know a priori that if she is thinking that *p*, then *E*.

The key difference between the two versions of the argument is that when the second premise is (2\*) the argument involves material implication instead of logical implication. (Aarnio, 436) Aarnio refers to both versions of this argument as ‘McKinsey-style Incompatibilist arguments’ since they take the form of the arguments McKinsey used in his papers.

Before moving to Aarnio’s central argument it is important to discuss an aspect of her argument that is necessary but the reason for its necessity is not easy to see until the conclusion of her argument. This aspect is that she will be using demonstratives within the propositions which require a referent mostly because demonstratives are the least controversial of terms to which content externalism is supposed to apply. As Aarnio states herself, “I take there to be very

strong support for the claim that demonstratives function as singular terms (not quantifiers) that rigidly designate an object with respect to all possible worlds, and that if a non-referring demonstrative term occurs in a sentence, then that sentence fails to express a proposition.”

(Aarnio 435) Essentially, demonstratives are the type of terms that it seems extremely implausible for them to be used without there being an object they refer to and thus unable to escape the grasp of content externalism. The only difference this means for us right now is that the above arguments will now take the following form:

(THAT1) Suzy can know a priori that she is thinking that that [Fred] is poisonous.

(THAT2) The proposition that Suzy is thinking that that [Fred] is poisonous logically implies the proposition that that [Fred] exists.  
Therefore,

(THAT3) Suzy can know a priori that that [Fred] exists.

In turn, the alternate version of the argument involves substituting the following premise for

(THAT2):

(THAT2\*) If Suzy is thinking that that [Fred] is poisonous then that [Fred] exists.  
(Aarnio, 435-6)

Aarnio’s central argument is that premise (THAT1) in both versions of the argument is unnecessary because the argument being setup as it is in both versions requires two assumptions for its validity. These assumptions are the validity of some closure principles and a priori knowledge of certain externalist theses. There are five closely related closure principles which play distinct roles that she utilizes throughout the paper. Although, two of them are key principles because not only does the validity of McKinsey-style arguments depend on the validity of those two principles but they also blaze the trail that leads to the necessity of the second assumption. These two key principles are as follows:

**Closure under a priori knowable logical implication (CAKL)**

Necessarily, for any person  $x$  and any propositions that  $p$  and that  $q$ , if  $x$  can know a priori that  $p$  and  $x$  can know a priori that the proposition that  $p$  logically implies the proposition that  $q$ , then  $x$  can know a priori that  $q$ .

**Closure under a priori knowable implication (CAK)**

Necessarily, for any person  $x$  and any propositions that  $p$  and that  $q$ , if  $x$  can know a priori that  $p$  and if  $x$  can know a priori that if  $p$  then  $q$ , then  $x$  can know a priori that  $q$ .

The closure principle (CAKL) is necessary for the validity of the McKinsey-style argument using premise (THAT2) because without the principle there is no reason to think that Suzy could know the consequent of the logical implication a priori, that that [Fred] exists, just because she knows the antecedent a priori, that Suzy is thinking that that [Fred] is poisonous. The validity of the argument depends on a principle which dictates that a priori knowledge of the antecedent of a logical implication leads to the possibility of a priori knowledge of its consequent. For the same reason, (CAK), or a principle which performs the same role as the one above but with material implication instead of logical implication, is necessary for the validity of a McKinsey-style argument using premise (THAT2\*).

However, in both situations above there is also the problem that it seems like both (THAT2) and (THAT2\*) must be knowable a priori for the arguments to be valid as well. Even with the closure principles it seems wrong to think that Suzy could know that that [Fred] exists a priori without her knowing a priori that her thought logically or materially implies it. The key here is that both principles involve the *knowability* of the existence of that [Fred] rather than claiming that the existence of that [Fred] will be known as long as the conditions are fulfilled. The proposition that that [Fred] exists is knowable a priori as long as Suzy's thought in (THAT1) is known a priori and the closure principle is valid but the existence of that [Fred] can't be

actually known a priori until the logical implication and the closure principle are both known a priori.

Now even though it seems like certain externalist theses must be knowable a priori in order for the McKinsey-style arguments to be valid, this isn't necessarily the case. This is the subject of her second section and it is rather significant because in her third section Aarnio provides good reasons for why Incompatibilists should not want to make such an assumption. Aarnio claims that there are two possible routes that the Incompatibilist can take to avoid this. Either they can argue that (THAT2) does not actually have to be knowable a priori or it is knowable a priori but that knowability is conditional on the privileged access premise. (Aarnio 437)

Aarnio argues that both routes rely on two other closure principles for their validity whose descriptions I will wait on providing till the elaboration of her argument for clarity's sake. Aarnio claims that any justification for those principles also justifies a 'weaker' principle which essentially states that logical implications are essentially logical truths which have the property of being logically implied by any proposition. (Aarnio 439) This would entail that as long as Suzy can know any proposition a priori, which Aarnio regards as a very plausible assumption, Suzy can also know any logical truth a priori as well. Aarnio then makes a crucial move by stating that as long as Suzy can know one proposition a priori she is then in a position to know (THAT2) a priori since it just describes a logical truth which is logically implied by the proposition. Furthermore, the logical implication in (THAT2) simply states that if Suzy is thinking that that [Fred] is poisonous then that [Fred] exists which is precisely the same conditional expressed by (THAT2\*). So by trying to avoid making the unconditional assumptions of certain externalist theses, the Incompatibilist was led back to a point where it was

shown that they must at least grant Suzy the capability of knowing an externalist thesis like (THAT2) and (THAT2\*) a priori.

At first this does not seem like a big problem, but Aarnio points out in her third section that any justification for a priori knowledge of (THAT2) and (THAT2\*) also supports a priori knowledge of two propositions which she labels as (\*) and (\*\*) which will be described later. However, as Aarnio argues, when you combine (THAT2) and (\*) with (CAKL) or (THAT2\*) and (\*\*) with (CAK) we are able to get to the conclusion (THAT3) without the use of the privileged access premise (THAT1).

Now that her central argument is on the table it should be easy to follow how the argument actually plays out in her description. While Aarnio does admit that there are arguments against closure principles which could be a big weak point in her argument, she claims that she only makes use of the principles that an Incompatibilist must appeal for their arguments to be valid so if there were a significant problem with one or all of the principles the Incompatibilist's argument would end up being invalid as well. So any instance of closure principles in this paper should only be questioned in considering their necessity to the Incompatibilist argument and not when it pertains to its independent validity.

In the second section, Aarnio states that there are two routes available to the Incompatibilist to avoid making the devastating assumptions. The first does not assume a priori knowledge of (THAT2) or other externalist theses that perform the same function while the second does assume such knowledge but makes it conditional on the privileged access premise. (Aarnio, 437) According to Aarnio, both routes each involve one of the following two closure principles:

**Closure of a priority under logical implication (CA)**

Necessarily, for any person  $x$  and any propositions that  $p$  and that  $q$ , if  $x$  can know a priori that  $p$ , and the proposition that  $p$  logically implies the proposition that  $q$ , then  $x$  can know a priori that  $q$ .

**Partial closure under logical implication (PLC)**

Necessarily, for any person  $x$  and any propositions that  $p$  and that  $q$ , if  $x$  can know a priori that  $p$ , and the proposition that  $p$  logically implies the proposition that  $q$ , then  $x$  can know a priori that if  $p$  then  $q$ .

The first route employs (CA) while the second utilizes (PLC) as well as (CAK). (Aarnio, 438) In order to show that neither route works out the way the Incompatibilist had hoped, Aarnio shows how an absurd conclusion can be generated using the second route and then explains that the same conclusion can be derived even more directly with the first route.

According to Aarnio, applying (PLC) to (THAT1)-(THAT3) makes the possibility of a priori knowledge of (THAT2\*) conditional on the privileged access premise being true. If we take (THAT1) to be  $p$  then as long as (THAT1) logically implies (THAT2\*) then by (PLC) Suzy can know (THAT2\*) a priori. Then through (CAK) and a priori knowledge of (THAT2\*) we can derive (THAT3). However, Aarnio wants to argue that the motivations for (PLC) also seem to be the same motivations for a different principle which allows us to get a priori knowledge of (THAT2\*) without the necessity of a priori knowledge of (THAT1) as long as Suzy is granted the possibility of a priori knowledge of at least one proposition which does not pertain to her mental states. This last requirement is necessary for proving that the a priori knowledge of (THAT2\*) is not conditional on the privileged access premise. She never labels the principle she refers to but we will label it (PLC\*) and it can be described as follows:

(PLC\*) For any propositions that  $p$  and that  $q$ , if the proposition that  $p$  logically implies the proposition that  $q$ , then a subject can know a priori that if  $p$  then  $q$ . (Aarnio 438)

It is easy to see that this principle shares the same amount of justification as (PLC) since all that is taken out is  $x$  being the subject and the proposition that  $x$  can know that  $p$  a priori. A subject's ability to know a priori 'that  $p$  logically implies that  $q$ ' is not diminished by the omission of the a priori knowledge of 'that  $p$ ' because a priori knowledge of the proposition 'that  $p$ ' does not entail a priori knowledge of everything the proposition logically implies. Essentially all (PLC\*) states is that all cases of logical implication have the possibility of being known a priori which is also implied by (PLC). All this revision does is clarify that the a priori knowledge of the proposition 'that  $p$ ' is not what allows for the possibility of a priori knowledge of what the proposition logically implies. Instead it highlights that simply being a logical implication is what allows for the possibility of a priori knowledge of the proposition 'that  $p$  logically implies that  $q$ '.

Aarnio is now able to show how this second route fails to accomplish what the Incompatibilist had hoped it would. The inclusion of (PLC\*) is meant to show that the nature of logical implication is that of a logical truth. The current prevailing doctrine concerning logical truths has it that a logical truth is logically implied by any true proposition. (Aarnio 439) So if we allow Suzy a priori knowledge of at least one proposition such as  $2+2=4$ , which we will label  $r$ , and we state that the proposition that  $s$  logically implies the proposition that  $u$ , then it must be the case that the proposition that  $r$  logically implies that (if  $s$  then  $u$ ). (Aarnio 439) Since we assumed that Suzy can know  $r$  a priori, by (PLC) Suzy can know a priori that if  $r$  then (if  $s$  then  $u$ ). Due to this case in particular Aarnio believes that (PLC) validates the following principle:

Necessarily, for any subject  $x$  and any propositions that  $s$  and that  $u$ , if there is any proposition  $r$  such that  $x$  can know a priori that  $r$ , and if the proposition that  $s$  logically implies the proposition that  $u$ , then  $x$  can know a priori that if  $r$  then (if  $s$  then  $u$ ).

This principle combined with (CAK) gives validation to the following argument:

Suzy can know a priori that  $r$ .

Suzy can know a priori that if  $r$  then (if  $s$  then  $u$ ).

Therefore,

Suzy can know a priori if  $s$  then  $u$ . (Aarnio 439)

When we substitute ' $r$ ' with ' $2+2=4$ ', ' $s$ ' with 'Suzy is thinking that that [Fred] is poisonous', and ' $u$ ' with 'that [Fred] exists' then we get the conclusion 'Suzy can know a priori that if Suzy is thinking that that [Fred] is poisonous, then that [Fred] exists' which is essentially (THAT2\*). (Aarnio 440) Aarnio asserts that the same conclusion can be derived even easier in the first route when using (CA). While she never actually elaborates on how the first route arrives at the same conclusion as the second, the most obvious reason for why the first arrives at the conclusion more directly is because the proposition that can be known a priori is the same as the antecedent proposition in the logical implication. The benefit of this is that the a priori knowledge of that  $p$  alone provides us with the proposition that is enough to logically imply the logical truth of if that  $p$  then that  $q$  which then gives us a priori knowledge of that  $q$  since we have the antecedent of both conditionals.

Aarnio sums up the second section through showing that the Incompatibilist fell back on (PLC) to avoid the unconditional assumption of a priori knowledge of (THAT2) or (THAT2\*) by making such assumed knowledge conditional on the privileged access premise. However, the conjunction of (PLC), (CAK), an externalist thesis similar to (THAT2), and the assumption that any subject in a McKinsey-style Incompatibilist argument can have a priori knowledge of at least one proposition not pertaining to their mental states results in the fact that the subject in question can have a priori knowledge of an externalist thesis similar to (THAT2\*). (Aarnio 440)



Aarnio's third section is where she aims to prove that the possibility of a priori knowledge of an externalist thesis similar to (THAT2\*) can result in a priori knowledge that [Fred] exists. This not only presents a problem for those who want to advocate that externalist theses like (THAT2\*) can be known a priori because Aarnio argues that such a priori knowledge allows us to arrive at a priori knowledge that [Fred] exists independent of the privileged access premise. This is where Aarnio's use of demonstratives presents a crucial move in her argument. The support for externalist theses like (THAT2) and (THAT2\*) comes from the object-dependence of their content. (Aarnio 441) Any instance of any person thinking such content is a case where the demonstrative *that* involved in the content must have an object which it refers to. While both theses differ in how strong the entailment is between an instance of thinking a thought about such externalist theses and the existence of the object it refers to, it is still the case that both theses do have externalist entailments. (Aarnio 441) Therefore, Aarnio argues, the following propositions are supported by the same things supporting (THAT2) and (THAT2\*) respectively:

(\*) (THAT2) logically implies the proposition that that [Fred] exists.

(\*\*) If (THAT2\*), then that [Fred] exists.

According to Aarnio, the problem for Incompatibilists is that there is no possible way to explain how we could allow someone a priori knowledge of (THAT2) or (THAT2\*) but restrict them from a priori knowledge of (\*) or (\*\*). (Aarnio 441) She claims that any a priori knowledge of (THAT2) or (THAT2\*) would have to be, "based on being able to recognize, a priori, that the proposition that that [Fred] is poisonous is object-dependent – whether this would involve recognizing that it is a Russelian proposition with Fred as a constituent, that it has an object-dependent sense with *that*<sub>FRED</sub> as a constituent, or something else." (Aarnio 441)

Although, the simple ability to know a priori that a proposition is object-dependent also brings with it the ability to know a priori that any distinct proposition which contains the former proposition will also be object-dependent. Since (THAT2) and (THAT2\*) are propositions that contain an object-dependent proposition within them, if we grant that Suzy can know (THAT2) and (\*) a priori then the following argument is valid through the application of (CAKL):

Suzy can know (THAT2) a priori.

Suzy can know a priori that (THAT2) logically implies the proposition that that [Fred] exists.

Therefore,  
Suzy can know a priori that that [Fred] exists.

In turn, if we grant Suzy a priori knowledge of (THAT2\*) and (\*\*) then the following argument is valid through an application of (CAK):

Suzy can know (THAT2\*) a priori.

Suzy can know a priori that if (THAT2\*), then that [Fred] exists.

Therefore,  
Suzy can know a priori that that [Fred] exists.

Aarnio concludes that any rationale which allows Suzy a priori knowledge of (THAT2) or (THAT2\*) also allows her knowledge of (\*) or (\*\*) which then allows Suzy to arrive at the same absurd conclusion of the McKinsey-style Incompatibilist arguments without the involvement of the privileged access premise. Though, she leaves it an open possibility that it may very well come to be that the absurd conclusion of the McKinsey argument is only generated when assuming both content externalism and privileged access even if the Incompatibilists were not justified in their use of McKinsey-style Incompatibilist arguments so far. Aarnio's argument is not infallible so it is only right that she admit to the possibility of error but as she herself states, "this places the ball squarely within the court of the Incompatibilist." (Aarnio 442)

In her conclusion, Aarnio makes it very easy for the reader to extract the moral of her paper. Apart from potential controversy surrounding the use of closure principles her argument leaves open four options. Either we reject externalism, deny that externalist theses like (THAT2) or (THAT2\*) can be known a priori, argue that such a priori knowledge is conditional on the privileged access premise, or accept the fact that we can have some a priori knowledge of the external world. (Aarnio 443) Aarnio herself thinks that the fourth option is the way to go but such a controversial claim should be left for another paper so that the arguments are not tarnished by its radical nature.

Unless one were to be persuaded that her use of closure principles or claim of the hold that content externalism has on demonstratives should be rejected, Aarnio's argument appears airtight. There is, however, a consequence of her argument that is easy to overlook which calls the classification of her argument as one of compatibility into question. This is because a new incompatibility argument can be extracted from her own argument which deals with a conflict between the possibility of a priori knowledge of externalism and the nature of demonstratives. If one were capable of having a priori knowledge of content externalism and the nature of demonstratives it would be possible for someone to have a thought using a demonstrative and immediately know a priori that the referent of the demonstrative exists. As Aarnio states in the beginning of her paper, any instance of a demonstrative is one where it acts as a rigid designator and any case of a sentence where the demonstrative in it does not have a referent is one where the sentence fails to express a proposition. In such a case, a sentence or thought that fails to express a proposition will not have any content and thus doesn't apply to content externalism so a priori knowledge of content externalism does not apply. However, in cases where there is content and content externalism does apply, a priori knowledge of content externalism and the

nature of demonstratives does leave open the possibility of a priori knowledge of contingent facts about our environment. This possibility alone presents an apparent incompatibility.

One way around this rejection is to adopt the common method used when the topic is natural kinds and claim that we are not able to know a priori if the demonstrative used in our thought refers to anything existing in the environment or not. While this is a good solution it overlooks the fact that taking this route severs the necessary tie between content externalism and demonstratives which is one of the least controversial of the few remaining types of terms to which content externalism is supposed to apply. While it remains true that referring demonstratives are still inescapably tied to content externalism, the possibility of non-referring demonstratives would create enough doubt that would prevent us from being able to know whether a particular instance of a demonstrative is referring or not. This is not a problem for the conclusion of Aarnio's argument because it allows for the compatibility of privileged access and content externalism since the absurd conclusions of McKinsey-style instance arguments can never be actualized without knowing whether the demonstratives used in them refer or not.

On the other hand, without the knowledge of whether a given demonstrative refers or not Aarnio cannot make her central argument altogether. It was the fact that we could know a priori that content externalism applies to a given demonstrative that allowed Aarnio to create the incompatibility independent of privileged access. Without this part of her argument she did not create a reason for why the incompatibilist would not want to make the assumptions described in the second sections because the assumptions no longer lead to an incompatibility independent of privileged access. So without the ability to know a priori whether a given demonstrative refers or not, the most Aarnio's argument can do is back an incompatibilist into a corner and force them to make some controversial assumptions. The problem is that those assumptions do not have

implications that overthrow the whole incompatibility argument anymore and the incompatibilist can just sit back in the corner they were backed into until Aarnio makes the next move. The ball is not squarely in the incompatibilist's corner anymore.

Even though Aarnio's theory has some serious implications, it should not be seen as a completely useless endeavor. Her method of trying to create incompatibility without privileged access was an original idea that had not been tried up until this point. In the end it may not have gotten the job done but it does not suffer from any fate different than any of the other compatibilist arguments which had been advances prior to her argument. The only theory which will not fail will be the one which solves the problem altogether such a solution is not likely to be found without the stepping stones like Aarnio's argument being put in place first.

## Chapter VIII

Like Aarnio, Andre Gallois also presents an objection targeted at Michael McKinsey's Incompatibilist argument. Although, unlike Aarnio who phrases the McKinsey argument in terms of the Twin Earth thought experiment, Gallois puts it in terms of modality. While there are many more premises in the modal form of the argument, there are not as many assumptions left unstated like there are in the version presented by Aarnio. The modal form of the McKinsey argument is as follows:

- (1)  $\Diamond (\exists x) (x \text{ knows a priori that } x \text{ thinks that } x \text{ lives in a watery world})$
- (2) ETC
- (3)  $\text{ETC} \rightarrow \Diamond (\exists x) (x \text{ knows a priori that ETC})$
- (4)  $\Diamond (\exists x) (x \text{ knows a priori that ETC})$
- (5)  $\text{ETC} \rightarrow \Box (x) (x \text{ thinks } x \text{ lives in a watery world} \rightarrow x \text{ lives in a watery world})$
- (6)  $\Diamond (\exists x) (x \text{ knows a priori that (5)})$
- (7)  $\Diamond (\exists x) (x \text{ knows a priori that } \Box (y) (y \text{ thinks that } y \text{ lives in a watery world} \rightarrow y \text{ lives in a watery world}))$
- (8)  $\Diamond (\exists x) (x \text{ knows a priori that } x \text{ lives in a watery world})$

Gallois labels this as the McKinsey\* argument because it is different from the argument McKinsey developed in that it does not make any assumption as to whether ETC involves conceptual implication or metaphysical entailment. (Gallois 158-159) If it wasn't clear already, ETC is an acronym which represents the theory of content externalism. Since the McKinsey\* argument does not make any controversial assumption on that topic it is more resilient to objections. The proposition (8) is equivalent to the negation of the third proposition of

McKinsey's triad which is in agreement with the conclusion commonly seen as absurd. Premise (2) is just the assumption that content externalism is true which is required for this to be an argument about the compatibility of content externalism with privileged access. Premise (5) simply describes the hold that content externalism has over our concepts which were obtained from our external environment. While there is much controversy surrounding the topic of which concepts are and are not affected by content externalism, the one fact that is not up for debate is that if a concept were acquired from interacting with a certain object in the external environment then the individual who possesses the concept must live in an environment containing at least one object to which that concept refers. However, it is one thing to assume ETC is true but it is a completely different thing to assume that it is knowable a priori which is what premise (3) claims possible. Many would argue that because ETC is a necessary truth it should be possible to come to know it a priori. Others like Saul Kripke would object to this argument by claiming that some necessary truths aren't knowable a priori. While it may be necessarily true that an individual with a concept obtained from an object in the environment must live in an environment containing that object, it is not necessarily the case that someone could come to know this truth completely independent of experience. Whether or not this is possible is a significant matter because if it turns out that (3) is false then (6) must also be false which would cause the whole argument to collapse.

One last thing must be assumed to validate the McKinsey\* argument. In tune with Aarnio, Gallois claims that a closure principle is required for this validation. This principle is the proposition that follows:

$$(9) \Diamond (\exists x) (x \text{ knows a priori that } (p \ \& \ p \rightarrow q)) \rightarrow \Diamond (\exists x) (x \text{ knows a priori that } q)$$

Gallois acknowledges that closure principles are controversial within this topic but he believes that the principle which is objected to most often makes much stronger claims than (9). The principle he is speaking of claims that if someone knows *p* and knows that *p* implies *q* then they also know *q*. (9) simply claims that if someone *could* know a priori that *p* as well as the fact that *p* implies *q* then they *could* know a priori that *q*. (Gallois 160) (9) makes no claim as to whether someone could come to know that *q* a priori in the actual world or even in possible worlds. It just states the conditions under which a priori knowledge of *q* is possible, not for the actualization of such knowledge. However, it is easy to see how such a closure principle is necessary for the validation of the McKinsey argument. Without the principle we would not be able to get (4) from (2) and (3), (7) from combining (4) and (6), or most importantly (8) by combining (1) and (7).

Gallois states that (3) and (6) along with (5) are the premises in the McKinsey argument that are the most vulnerable to objection. (Gallois 186) Even though I characterized (5) as expressing a claim of content externalism which is not up for debate, Gallois's reason for believing it is vulnerable has to do with a lack of elaboration about what kind of 'water' concept is involved. While (5) may express a necessary truth for natural kind concepts, when it comes to descriptivist concepts (5) is completely wrong. Gallois's thesis is that even if we grant the Incompatibilist (3) and (6), our inability to know a priori whether the concept used in our thought is a natural kind or descriptivist concept leads to necessary revisions in many of the premises. After a couple revisions, the new version of (8) will no longer be an absurd result.

Gallois starts off by considering a point made by Anthony Brueckner in his direct response to McKinsey's original argument. We cannot blindly accept (5) because it is possible to imagine a scenario where someone can have a thought about a natural kind which does not exist in the individual's environment if they had gone through sufficient theorizing about the



properties of that natural kind if it had existed. (Gallois 183-184) For instance, it is possible to imagine someone skilled in atomic chemistry that lives in a world with both oxygen and hydrogen atoms but both atoms have never actually combined. In this world, this person could potentially theorize about the substance that would result from combining two hydrogen atoms with one oxygen atom. This possible situation leads to (5) being changed to:

$$(5') \quad \text{ETC} \rightarrow \Box (x) (x \text{ thinks } x \text{ lives in a watery world} \rightarrow (\text{no one has theorized about water} \rightarrow x \text{ lives in a watery world}))$$

This revision also changes (7) to include that no one has theorized about water in the conditional but more notably it changes (8) to:

$$(8') \quad \Diamond (\exists x) (x \text{ knows a priori that } (\text{no one has theorized about water} \rightarrow x \text{ lives in a watery world}))$$

While (8') is not as counterintuitive as (8), Gallois states that it is counterintuitive enough to be rejected and thus requires that we reject one of the premises that led to its derivation because without such a rejection we would be forced to give up either content externalism or privileged access.

Naturally Gallois decides to reject the weakest of the premises which would be (6') because the addition of the theorizing condition causes (6') to have a weaker claim than (6). Gallois uses a thought experiment involving phlogiston as justification for this rejection. The concept of phlogiston seems to function in a similar way as the natural kind concept of water by picking out a specific substance, except that the substance that the concept of phlogiston refers to does not actually exist. However, if it turned out that we were wrong about oxidation and the substance phlogiston actually existed then it would follow that phlogiston is a natural kind and

any concept derived from interacting with it would be a natural kind concept. Therefore, Gallois claims that in this case ETC implies the following:

- (10) (x) (x thinks that phlogiston was believed in by eighteenth-century chemists  $\rightarrow$  (no one has theorized about phlogiston  $\rightarrow$  phlogiston exists)).

The argument, albeit a bad one as Gallois points out, is that if we grant an individual a priori knowledge that they are having a thought about phlogiston and also assume that ETC is true then we could know a priori that if no one has theorized about phlogiston then phlogiston exists.

Disregarding the fact that there are three big 'ifs' that must be satisfied prior to reaching a priori knowledge that phlogiston exists, it is a bad argument because it is easy to imagine a counterfactual situation where someone has a thought about phlogiston and the eighteenth century chemists never theorized about it but yet it still doesn't exist. Chances are that the first chemist who theorized about phlogiston had some thought involving the substance before he actually began theorizing about it. While the concept of the substance used in that thought would have been a very underdeveloped concept, it nonetheless was present in the thought and attempted to refer to a natural substance that the chemist believed to exist. According to Gallois, this argument goes awry in the assumption that ETC implies (10) at all times. He claims that ETC would only imply (10) if the concept of phlogiston used in the thought were a natural kind concept. Whether this claim is accurate is something that will be examined shortly, but for the time being Gallois asks us to assume that ETC does in fact imply (10) when phlogiston refers to a natural kind. (Gallois 185-186)

Here is where Gallois presents what he labels the Semantic Variation thesis. When a thought involves a concept, this concept will function differently within the thought depending on whether the concept refers to a natural kind, does not refer to anything, or refers to anything

that is not a natural kind. If the concept refers to a natural kind then it will perform as a rigid designator where only things that are a member of that natural kind can be the referents of that concept. In turn, if the concept does not refer to any existing things then it is a descriptivist concept which performs as a non-rigid designator and anything which meets the description set for it will be a referent of the concept. For phlogiston, this description would be ‘the stuff thought by some eighteenth-century chemists to have negative weight’. (Gallois 186)

Gallois now evokes the question concerning whether ETC implies (10) at any time and if so under which conditions does ETC imply (10)? According to (10), the occurrence of a thought in someone that phlogiston was believed in by eighteenth-century chemists entails that, if no one has theorized about phlogiston, phlogiston exists. Gallois labels this sentence which expresses the proposition in (10) as S. Now Gallois also takes the two different propositions expressed by S when phlogiston acts as a natural kind or a descriptivist concept and labeled them as P1 and P2 respectively. According to Gallois, the only way that an individual could be capable of answering the question of whether ETC implies (10) is if we know whether S expresses P1 or P2. Since it was already assumed for the time being that ETC implies (10) when S expressed P1, Gallois examines what (10) would look like when S expresses P2.

(10\*) (x) (x thinks that phlogiston was believed in by eighteenth-century chemists → (no one has theorized about the stuff thought by eighteenth-century chemists to have negative weight → the stuff thought by eighteenth-century chemists to have negative weight exists)).

Gallois claims that (10\*) clearly does not follow from ETC. When phlogiston acts as a non-rigid designator then anything which meets the description would be a referent of the term. One possible objection to this claim that (10\*) does not follow from ETC is that phlogiston had been theorized about by eighteenth century scientists which is why (10\*) does not necessitate its

existence. However, the reason this objection does not present a problem is because we can imagine a counterfactual situation where phlogiston had not been theorized about by eighteenth century scientists. In such a case, the subject's thought about phlogiston as a descriptivist concept would, by (10\*), necessitate that a referent of phlogiston exists since it had not been theorized about in the past despite the fact that the concept was not obtained from any referent. Clearly this consequence is absurd because if it were impossible for a subject to have a thought about a descriptivist concept that had not been previously theorized about then it would be impossible for there to be any descriptivist concepts at all. Surely theorizers must be able have thoughts about the concepts they are currently theorizing prior to the completion of their theorization. Therefore, Gallois concludes that ETC does not imply (10) when phlogiston is a descriptivist concept.

Due to this thought experiment Gallois believes the following claim should be supplemented into the Semantic Variation thesis. "Since the semantic function of 'phlogiston' depends, in part, on the existence of phlogiston as a natural kind, determining the semantic function of 'phlogiston' requires empirical investigation. However, one needs to determine the semantic function of 'phlogiston' in order to tell whether (10) is identical with something implied by ETC. Hence, no one can know a priori that ETC implies (10). Even if ETC can be known a priori, there is no reason to think that (10) can." (Gallois 187) Essentially what this all boils down to is that even if we assume a priori knowledge of ETC and of a thought involving a certain concept, we are still not capable of obtaining a priori knowledge that what the concept refers to actually exists in the external environment without a priori knowledge of whether the concept involved is a natural kind concept or not. Since knowledge of whether it is a natural kind

concept or not is only obtainable through empirical investigation, it is impossible for anyone to know a priori that the referent of a concept used in their thought actually exists.

According to Gallois, (5') is only necessarily true when 'water' refers to a natural kind. When 'water' performs as a non-rigid designator it is up in the air whether the conditions for (5') being true are satisfied. Due to this fact it is no longer the case that (5') can enjoy the privileged status of being a priori knowable simply by being a necessary truth. The argument in the last paragraph proved that whether or not a concept refers to a natural kind is not knowledge that can be acquired a priori and since this a priori knowledge is necessary for a priori knowledge of the truth of (5'), (6') is necessarily false.

Gallois considers an objection to his semantic variation thesis. Even though we cannot come to know (6') a priori because we cannot know a priori whether the concept used in our occurring thought refers to a natural kind or not, the problem at hand is to try and find a way to describe how we can have a priori knowledge of our occurring thoughts in the first place. Before we can ask the question of how we can have a priori knowledge of the status of a given concept used in a thought we need to first ask if we can to any extent have a priori knowledge of the content of the occurring thought. If we can't have a priori knowledge of any part of the content of the thought then we couldn't have a concept before us whose status as a natural kind or a descriptivist can be questioned.

In response to this objection, Gallois asks us to recall his argument against Boghossian. Even though someone may not be able to know their thought is about water rather than twater, they may still be able to know a priori their thought that water is wet because they can contrast this occurring thought against the potential, relevant alternatives that they do know they are not

entertaining at the moment. Similarly with the semantic variation thesis, when it comes to thoughts about water there will be a pair of propositions that could be expressed by those thoughts that an individual may not be able to distinguish between a priori. However, Gallois argues that this does not mean that no one can know a priori the content of their water thought. For that to be true, the proposition that would be expressed by the thought 'Water is wet' if water didn't exist would have to be included in the set of relevant alternatives for the proposition that water is wet. (Gallois 188) Gallois never clearly states how this overcomes the objection but I think what he is getting at is that the proposition expressed by 'Water is wet' if water didn't exist would be a proposition utilizing a descriptivist concept and such a proposition could not belong to the set of relevant alternatives for propositions utilizing a natural kind concept. Water concepts that are non-rigid designators are not relevant alternatives to ones that are rigid designators because they could never be interchangeable regardless of whether the world is one where water exists or doesn't exist. What I mean by interchangeable is that a proposition that expresses 'Water is wet' when utilizing a descriptivist concept will never express the same thing as a proposition that expresses 'Water is wet' utilizing a natural kind concept. While there are many different things that can be expressed by both propositions, the only difference necessary to justify this claim is that one implicitly expresses that water exists and the other doesn't.

Gallois evokes another thought experiment to show that it is unnecessary to prove that the pair of propositions discussed above do not belong to the same set of relevant alternatives in order to show that (6') is false. First, Gallois starts by claiming that even if the semantic variation thesis were false there is no reason to believe (6') is true. We are asked to imagine a world like ours where Albertine's concept of water is acquired from either direct or indirect contact with the substance. We must then ask the question of whether Albertine could have thoughts about water

if water did not exist and we are drawn back to the usual response that she could not unless she has theorized about water sufficiently. This leads to the postulation of (12):

(12) If Albertine has acquired the concept of water from interacting with water then the following conditional is necessarily true. If Albertine thinks she lives in a watery world, and no one has theorized about water, Albertine lives in a watery world.

Gallois argues that (12) has it that if Albertine has acquiring the concept of water from interacting with water then there is no way she could have water thoughts without living in a world with either water or water theorizers. However, (12) does not entail that there is no way Albertine can ever have thoughts about water unless she lives in a world without water or water theorizers. (Gallois 188)

To illustrate this point Gallois presents two different situations and in both it is open for questions in Albertine's eyes whether water exists or does not in her world. In the first situation Albertine sees an actual body of water and decides to use the term 'water' as a synonym for a non-rigid designator describing an odorless, colorless, tasteless liquid despite the fact that water exists and is a natural kind. If she asks herself whether she could entertain a thought about this concept of water in a world where water didn't exist and in this situation the answer would be yes because she could have had a thought about an odorless, colorless, tasteless liquid in a world without water. In this situation (12) would be false because the antecedent was satisfied but not the consequent.

In the second situation, Albertine acquires her concept of water in a way that ensures (12) is true. Since this could have been by either interacting with water or inhabiting a world with water theorizers, she is still constrained to asking herself the question of whether water exists or not in her world. Albertine then asks herself, if she had acquired the concept from interacting

with water could she have entertained thoughts about water in a waterless world? Here the answer would be no because we are assuming (12) is true and since we are assuming the antecedent would be satisfied from the direct interaction with water then water exists. However, if Albertine proceeds to ask herself in this situation whether she could have thoughts about water if it in fact did not exist then Albertine can in fact answer yes because the antecedent of the condition in (12) is false so her potential thoughts are not constrained by it.

Gallois claims that he has argued that the usual thought experiments which have been used to justify ETC do not support anything stronger than (12). (Gallois 189-190) What can be derived from those thought experiments is, “Someone who has acquired a natural kind concept C from interacting with a natural kind K could not have had C unless that natural kind exists.” (Gallois 190) Although, even if we agree with this claim we are not forced to conclude that no one could have C unless K exists. It is only when C was acquired from interacting with K that it is impossible to have C without K existing. If C was acquired in some other way then it is consistent with (12) and therefore ETC that we can have C without K necessarily existing.

Given this conclusion, we are forced to revise (5') and replace it with:

(5\*)  $ETC \rightarrow \Box (x) (x \text{ has acquired the concept of water from interacting with water} \rightarrow \Box (x \text{ thinks } x \text{ lives in a watery world} \rightarrow (\text{no one has theorized about water} \rightarrow x \text{ lives in a watery world})))$ .

Which then leads to (6') and (7') being revised and replaced with:

(6\*)  $\Diamond (\exists x) (x \text{ knows a priori that } (5*))$ .

(7\*)  $\Diamond (\exists x) (x \text{ knows a priori that } (y) (y \text{ has acquired the concept of water from interacting with water} \rightarrow (y \text{ thinks } y \text{ lives in a watery world} \rightarrow (\text{no one has theorized about water} \rightarrow y \text{ lives in a watery world}))))$ .



Gallois claims that we would be very wrong to deduce (8') from (1) and (7\*) and the most that can be deduced from (1) and (7\*) is this:

(8\*)  $\Diamond (\exists x) (x \text{ knows a priori that } (x \text{ has acquired the concept of water from interacting with water} \rightarrow (\text{no one has theorized about water} \rightarrow x \text{ lives in a watery world})))$ .

Gallois argues that (8\*) is a harmless conclusion that is far from being absurd. The McKinsey\* argument no longer presents an incompatibility between ETC and privileged access because many people would agree that we can know a priori that if we acquired our concept of water from interacting with water, and no one has theorized about water, then water exists.

In response to Gallois's overall argument, he does not leave open much room for objection. Apart from rejecting the use of a closure principle, the only other vulnerable aspect of his argument is the claim that we are able to know a proposition expressed by our thought a priori even if we are not able to contrast this proposition from at least one particular possible alternative proposition that could have been expressed by the thought. While it makes rational sense that we should be able to contrast a 'water' thought from a 'grass' thought even if we are not able to contrast a 'water' thought from a 'twater' thought, one could easily argue that in order to claim that a thought is *known* to be a 'water' thought one must be able to distinguish from all relevant alternatives and not just all but one. Postulation of such an argument would not completely defeat Gallois's compatibility argument because Gallois can always fall back on Falvey and Owens' argument that we can have introspective knowledge of content even without introspective knowledge of comparative content. The only part of Gallois's argument that would be lost by having to resort to Falvey and Owens' argument is that we could no longer claim that we can know a priori that our thought is a 'water' thought as opposed to a 'grass' thought. The

reason this would be lost is because Falvey and Owens claim that we are restricted from all introspective knowledge of comparative content where Gallois only claims that we are restricted from comparative content for corresponding concepts developed in unique scenarios like the Twin Earth switching situation. While losing this part of the argument would not be the end of the world for compatibilism, it does run counter to our intuitions of the types of things in our minds that we are supposed to have privileged access to.

Putting aside this potential vulnerability, there is one aspect of Gallois's theory which is extremely significant for the compatibilist cause. This aspect is the idea that we are not capable of knowing a priori whether the concept used in our thoughts is a natural kind or a descriptivist concept. While this is a very simple notion, it almost entirely explains why the absurd conclusion of McKinsey-style arguments can be generated in the first place. Surely it would be possible for someone to reach a priori knowledge that some particular object exists in the external environment when the person is armed with a priori knowledge of content externalism, the content of their thought, and that a concept within that content is a natural kind concept. However, by taking out a priori knowledge of the latter the possibility of a priori knowledge of the existence of some object is erased. As long as it is the case that content externalism does not necessarily apply to descriptivist concepts then we could never reach a priori knowledge that the referent of a descriptivist concept exists precisely because we can't be sure that the concept even has a referent without empirical investigation. This aspect of Gallois's argument allows us to have a priori knowledge of all of the content of our thoughts except for the status of the concepts involved as natural kind or descriptivist while maintaining content externalism.

It would appear that this alone could solve the entire compatibility debate but something is still missing. Despite the rationality in the idea of not being able to know a priori whether a

concept is a natural kind or descriptivist concept, we often do know that the concepts used in our thoughts do have referents because at some time prior to the thought we have interacted with the objects that helped shaped our concepts. A compatibilist would usually respond to this claim by postulating that this knowledge is a posteriori because it was obtained through empirical investigation but the fact of the matter is that upon instantiation of a thought we have this knowledge without performing any empirical investigation after the thought's instantiation. This would be another potentially absurd conclusion for incompatibilists to latch onto unless there were a way to describe the situation so that a posteriori knowledge can be part of knowledge obtained a priori in such a manner that it could not be considered absurd. Such a description is available and will be brought to light towards the end of Sarah Sawyer's chapter. Even though she only briefly mentions the issue of natural kind and descriptivist concepts, I believe she should have spent more time with it because further elaboration on its affect within her theory would make her theory stronger against objections than it originally appeared to be.

## Chapter IX

Sarah Sawyer starts off her paper by explaining that her motivation for writing it stems from a sudden realization that her theory invites a consequence that could potentially be used to refute her theory altogether. Although, instead of taking the usual route of amending the theory so that the consequence can be avoided, she believes the best path to take is to embrace the consequence. While it may not sound like an intelligent decision to embrace something considered absurd by many respected philosophers, she believes that further clarification pertaining to how this consequence is usually reached and to the descriptions of the processes involved will free this consequence from the label of absurdness.

This particular problem originates with the question of, “how can semantic externalism account for privileged access to the content of one’s thoughts?” (Sawyer 523) According to Sawyer, this question presents us with two problems where the second is a result of a solution to the first. The first problem starts with the fact that due to semantic externalism the set of possible and actual thoughts that a person may have is determined by the relations that the person stands in to their environment. The relations that the person stands in with either their environment or other people in their environment with whom they communicate are what allow or restrict the thoughts they potentially can and cannot entertain. “Given that a subject is unable to distinguish the specific environmental features upon which her thoughts essentially depend from other possible features upon which her counterfactual thoughts might have depended, how can a subject generally know what she thinks?” (Sawyer 523) Such knowledge would require empirical information about the relations they stand in with their own environment and the necessity of such knowledge seems to refute the principle of privileged access. It also denies the

first-person authority aspect of privileged access because it is possible that some other person could be in as good of a position as we are to know the content of our own thoughts.

According to Sawyer, a sufficient solution to this problem has already been discovered. The idea of second-order states discussed by Burge, among others, and attributed to Davidson holds the key to solving this problem. The same factors which determine the content of our first-order thoughts are also what determine the content of our second-order thoughts. She points out in a footnote that most philosophers who adopt this stance usually subscribe to some kind of 'containment principle' where the content of the first-order thought is embedded in the second-order belief but Sawyer herself does not want to be committed to any such principle; she just claims that the content between levels of thought will co-vary alongside one another. Since the same set of concepts are available to all our levels of thought there is no reason the content of a second-order belief cannot have the same content as the first-order thoughts. (Sawyer 524)

Although Sawyer never explains further, I assume that the reason she believes this solution overcomes the problem is because these second-order thoughts are reflexive and even though we cannot have privileged access to the content of our first-order thoughts we can access the content of the second-order thoughts. Usually this is due to the fact that the second-order thoughts will be self-verifying since their content is drawn from the first-order thought and can be known a priori because they are necessarily true due to the fact that you cannot have the second-order thought unless the first-order thought occurred.

Sawyer argues that the second problem arises from the alleged absurdity of the conclusion in the previous argument and is directed to the semantic externalist. Instead of the burden being placed on privileged access to prove its truth in light of semantic externalism,

semantic externalism must now prove its truth in light of privileged access. The apparent incompatibility has once again resurfaced with the following claim:

(EC) x could not have non-empirical knowledge of contingent facts about her environment.

(EC) expresses a claim about a priori knowledge that is considered by many philosophers to be necessarily true. Although he is a compatibilist, we saw in Chapter III that Anthony Brueckner concedes prior to developing his own argument that in the original McKinsey Argument the proposition that Oscar can know E a priori “embodies a claim about *a priori* knowledge which is obviously false on anyone’s view.” (Brueckner 197) E represents some contingent fact of the environment that Oscar’s ability to think his thought ‘Water is wet’ depends on due to semantic externalism. Sawyer constructed (EC) so that its negation expresses the claim that Brueckner conceded to be obviously false on anyone’s view. (Sawyer 524) By constructing (EC) to express the negation of the claim that Brueckner states would be obviously false on anyone’s view, Sawyer is trying to make us feel safe in making the assumption that Brueckner would also concede that (EC) expresses a claim that would be necessarily true on anyone’s view including compatibilists. That (EC) is necessarily true is the claim that Sawyer wishes to challenge.

The problem then is that through semantic externalism we can know the necessary conditions that must be met to possess any particular concept and if we assume that we have privileged access to the content of our thoughts as well as our concepts then we are able to infer a priori knowledge of many things we should only be able to know empirically. Since this knowledge was acquired through introspection and conceptual analysis alone, this argument illustrates yet another way in which semantic externalism and privileged access are apparently incompatible. Sawyer labels this the *Argument from Privileged Access*. (Sawyer 524)

According to Sawyer, the main reason why the Argument from Privileged Access leads to the alleged incompatibility is because it allows for the generation of ‘instance arguments’ which Sawyer labels as such because they allow someone to go from the instance of a thought to contingent facts about the world. (Sawyer 524-525) Instance arguments usually take the following form:

- (1) I am thinking a *water*-thought
- (2) If I am thinking a *water*-thought, then I’m in a water-world
- (3) Therefore, I am in a water-world

It is crucial to notice that for Sawyer, as many philosophers would agree, that for a subject to be in a water-world is not necessarily for them to be in a world containing water as a natural kind. The subject may be in a water-world which does not contain any water as a natural kind as long as the subject is part of a community which has the concept of water. Sawyer alongside Andre Gallois also points out that knowledge of the status of a concept, as a natural kind or descriptivist concept, is not required for knowledge of the possession of a concept. However, this fact is not as significant for Sawyer’s argument as it was for Gallois’s argument.

Through introspection we arrive at premise (1) and by conceptual analysis we get premise (2). Premise (3) follows directly from (1) and (2) so it seems that we can have non-empirical knowledge that we live in a water-world but if this were possible then (EC) would be false. Since many consider the negation of (EC) to be absurd then it appears that we have arrived at a crossroads. Either we abandon semantic externalism or privileged access. Sawyer, on the other hand, takes a route that most philosophers never even considered and instead of going left or right she decides to turn around and go down the path we came from. She decides to challenge the absurdness of (EC).

Responses to the Argument from Privileged Access will fit into one of two categories. Those who are advocates of semantic internalism will respond by claiming that the argument is watertight and thus provides a reductio of semantic externalism. Others who do not want to give up semantic externalism will respond with arguments that attempt to find a flaw in the Argument from Privileged Access so that the conjunction of (1) and (2) does not necessarily entail (3). The former are considered Incompatibilists and the latter are considered Compatibilists. (Sawyer 525)

Sawyer's theory is rather unique in that she agrees with incompatibilists that the Argument from Privileged Access does lead to the possible derivation of empirical knowledge from introspection and conceptual analysis alone but her theory claims that such a derivation does not make semantic externalism and privileged access inherently incompatible. Due to this uniqueness Sawyer believes that it is essential to discuss two traditional compatibilist responses to the Argument from Privileged Access before she presents her own theory. In order to see why such an extreme method of argument, i.e. embracing an absurd conclusion, must be used we should learn why the more conventional tactics were destined to fail. However, first Sawyer's theory commits her to explain a crucial aspect of the principle of Privileged Access she is utilizing.

The idea of any sort of infallibility when it comes to self-knowledge has been a very unpopular and controversial topic since the inherent problems in the Cartesian view of the mind were pointed out. It would therefore seem like a risky move on Sawyer's part to adopt any sort of infallibility for the version of Privileged Access she wishes to use. However, she does in fact adopt the idea that an individual has infallible access to their own concepts and presents an argument for why this is the case which appears to be iron clad. First, we must look at the process of an individual attributing a thought to a subject. The first condition that must be satisfied before a correct attribution is possible is not the ability to make a correct inference from



the subject's actions but instead that the individual possesses the same concepts as the subject. (Sawyer 526) In the Twin Earth scenario, neither twin could possibly attribute their respective 'water' thoughts to the other because they do not possess the same *water* concepts. The attribution of a thought requires at the very least the ability to entertain the same thought yourself. As philosophers, we are not restricted as such because it is assumed that we possess the concepts discussed in our hypothetical thought experiments. Although, we are constrained to having to be more careful in our attributions since a correct attribution on our part also depends on our recognition of the environment our subject inhabits at the time of the attribution. The subjects themselves are not similarly constrained because they, as well as everyone else who inhabits the same environment, are "locked in" to the same way of thinking.

With all this on the table it's time to think about how we attribute thoughts to ourselves. The first condition here is already satisfied because there is no question that I possess the same concepts as the person thinking the thought because I *am* the person thinking the thought. Sawyer points out that this would explain the asymmetry in our ability to attribute a particular thought to ourselves and our ability to do the same with others even when the same exact behavioral evidence is available in both situations. However, Sawyer wants to go even further and claim that when we ascribe a given thought to ourselves it is not the case that we can be wrong in the ascription because the very act of ascribing the thought to ourselves involves thinking the thought as well. Even if the thought we had been thinking which motivated us to attribute some thought to ourselves in the first place were different than the thought we ended up ascribing to ourselves, the ascription itself could not be wrong because we ended up thinking the ascribed thought anyways. Sawyer's notion of 'thinks' is such that whenever the proposition 'S \*'s that p' is satisfied where '\*' is some form of propositional attitude, this satisfaction

necessarily entails that 'S thinks that p'. "The question *whether or not* you are entertaining a given thought is a mistaken question. The very formulation of such a question determines that the answer must be affirmative. Thoughts about the possession of a concept are self-verifying." (Sawyer 527) This leads Sawyer to the postulation of (PA) which she describes as the principle of privileged access with respect to concepts.

(PA) For all  $x$ , if  $x$  thinks she thinks that  $p$ , where concept  $F$  is an essential component of the thought that  $p$ ,  $x$  has the concept  $F$ .

Sawyer claims that under (PA) it would be impossible for even Davidson's swampman to be mistaken in thinking that he possessed some concept because he could not think a thought using a concept he did not possess.

Sawyer acknowledges that her account of privileged access is far from being complete but (PA) still serves the purposes it was intended for. The first purpose is that she claims that regardless of whatever potential problems privileged access and semantic externalism have on their own, the addition of semantic externalism to (PA) would not entail any *additional* problems. Sawyer never elaborates as to why this is the case but it would seem that what she has in mind is that semantic externalism provides us with knowledge of the necessary conditions for possessing a given concept and (PA) provides us with knowledge of the necessary conditions for thinking that we have a given thought. The simple, straightforward combination of knowledge of the necessary conditions for those two things is not enough to generate any new problems and therefore not inherently incompatible. The second purpose fulfilled by (PA) is that it is sufficient to generate instance arguments and thus susceptible to the Argument of Privileged Access without the necessity of complete knowledge of the current, conscious psychological state we are in. Sawyer states in a footnote that this means that instance arguments and the Argument from

Privileged Access are not weakened in cases where a mental state of jealousy is mistaken for hatred. (Sawyer 527)

Now that we have (PA) on the table it is time to look at the two traditional compatibilist responses that Sawyer wants to reject. The first response is to deny that we have knowledge of our concepts. It is tough to say how one could make this claim and remain a compatibilist because I cannot imagine how an argument would go that we have a priori knowledge of the content of our thoughts but we do not have knowledge of the concepts which are part of that content. We are assuming that full knowledge of the content of the thought is not required for us to know the thought in general but it is hard to see how any sense can be made of the content of the thought minus the content provided by the concepts and therefore how any knowledge can result from it. So the only way to make any sense of the claim is to adopt the view of Wittgenstein and argue that it doesn't make any sense to say that a proposition is known if it is guaranteed to be true. (Sawyer 527-528) On my interpretation of Wittgenstein's view, when something is necessarily true is it not that we come to learn it as we do contingent facts but rather we become *aware* of it as a necessary truth. I can't think of any other way to describe it because surely Wittgenstein isn't suggesting that our minds have an inherent restriction for processing information about necessary truths. Though I can see how we wouldn't want to call it knowledge because knowledge is learned and in order for something to be learned there must be a possibility of error; there must be the possibility that the fact I learn could have been different but I went through a process of ruling out those other potential alternatives. For necessary truths like  $1+1=2$ , we do not learn it but are instead made aware of its existence as a necessary truth. Since there is no possibility of thinking that we have a thought using concept F, when F is a concept we do not have, then our 'knowledge' that we have concept F cannot be considered knowledge. This

would solve the problem because we could never have knowledge of (3) let alone a priori knowledge of it.

The other response pertains to Nozick's conditional theory of knowledge and more specifically about its consequence that knowledge is not closed under known entailment. A common example is the following:

- (1') I am working at my desk
- (2') If I am working at my desk then I am not a brain-in-a-vat
- (3') Therefore, I am not a brain-in-a-vat

It would appear that there is no way that we could know (1') and (2') without that yielding knowledge of (3') but when the idea of possible worlds is introduced nothing is certain. Nozick's conditional theory of knowledge allows for the possibility that (1') and (2') could be known without entailing that (3') can be known because it is possible that each proposition could be assessed by different sets of possible worlds. (Sawyer 528) In Nozick's conditional theory a subject, S, can be said to know a proposition, P, if and only if 4 conditions are met: P is true, S believes that P, if P were false then S would not believe that P, and if P were true then S would believe that P. The only instances where S could know (1) and (2) without that resulting in knowledge of (3) are when the last two conditions are assessed for (1) and (2) on one possible world while those conditions are assessed for (3) on some separate possible world.

Sawyer wants to maintain that we shouldn't take this argument as a reason to think that knowledge isn't closed under known entailment. Just because we *could* assess the premises and conclusion relative to different sets of possible worlds does not mean that we *should* assess them in that manner. Moreover, we have more reasons to think that we should not assess them in that

manner than we have for thinking that we should. There is a possible world where (3) does not meet the 3<sup>rd</sup> condition because it is possible to imagine a world where S is a brain-in-a-vat but does not believe they are a brain-in-a-vat. However, in that possible world S could not know (2) either because in that world we could say with consistency that if S does not believe that they are a brain-in-a-vat despite being a brain-in-a-vat then it is possible that S also believes that they are working at their desk even though they are actually a brain-in-a-vat. This possibility alone invalidates the conditional in (2). If every world where both (1) and (2) are known is one where (3) is also known and every world where (3) is not known is one where either (1) or (2) is not known then there seems to be very good reasons for believing that we should not assess propositions which are paired together according to different sets of possible worlds.

The whole concept of possible worlds implies that each distinct possible world is different from any other in some way. The truth of any proposition could be affected by even the slightest change in circumstances so when we assess paired propositions we should not do so according to different sets of possible worlds. Upon moving from one possible world to another a proposition we previously considered to be true could become false. If two or more propositions cannot all be true on the same possible world then there is no justified reason to believe that we should hold them all to be true just because each one is true on distinct possible worlds. If we were to allow for this line of reasoning then the propositions 'I am currently in New York' and 'I am currently in California' could potentially both be considered true and, even worse, known. Under Nozick's conditional theory, the first proposition could be known because it meets the conditions when assessed in our actual world while the second proposition could also be known because it meets the conditions when assessed in some other possible world where I happen to be on vacation this week. The possibility that I can know that I am currently in New York and that I

am currently in California at the same time appears as absurd, if not more than, the conclusion of the McKinsey argument that it is possible for us to have a priori knowledge of the external environment. Considering that Sawyer is challenging the absurdity of the conclusion of the McKinsey argument, we should not conclude that due to this apparent absurdity we should not assess propositions according to different sets of possible worlds under Nozick's conditional theory. However, it is enough reason to claim that the closure principle which was under scrutiny by this use of Nozick's conditional theory should at least be safe for the purposes of her argument.

While there are other problems with compatibilist responses like those that deny knowledge of concepts or reject certain closure principles, the main factor that Sawyer wants to point out for why these arguments are unsatisfactory is that claiming that the conclusion of an instance argument is not *known* does not evade the Argument from Privileged Access. (Sawyer 528) The only people who were concerned about the Argument from Privileged Access but are relieved by the claim that the conclusion of an instance argument is not known are incompatibilists who do not want to sacrifice semantic externalism. This is because semantic externalism was backed into a corner and Wittgenstein and Nozick's responses gave it an out. They would be happy to sacrifice privileged access instead of semantic externalism. Compatibilists, on the other hand, would be far from relieved from Wittgenstein and Nozick's responses because it diminishes privileged access down to something a lot weaker than the principle we usually allude to in this debate and whose development was motivated by our intuitions about our access to our own minds. Furthermore, it would be extremely hard to persuade someone that they cannot know the conclusion of a type of argument that they have used many times before and has systemically allowed them to form true beliefs about their

environment. Sawyer wants to go as far as to claim that we can *know* that this method provides us with true beliefs. (Sawyer 528) If the Argument from Privileged Access depended on the question of whether or not the conclusions of instance arguments could be known then it would be pushed aside until that question is answered; but the real force of the argument is that in actuality people have the ability to systematically use instance arguments to obtain true beliefs about the environment and any theory which cannot account for how or why this method is capable of systematically yielding true beliefs is going to be found lacking something if it is not altogether wrong.

In the end, Sawyer's response to objections against the Argument from Privileged Access comes down to the fact that throughout human history we have used instance arguments to systematically arrive at true beliefs. Systematic generation of truth would seem to be a strict definition of what should be considered a reliable method and to date the best way to deem that we have epistemic warrant for holding a belief is that we arrived at this belief through a reliable method. (Sawyer 528) For this reason Sawyer believes that criticizing instance arguments is not the correct way to argue for compatibilism. Instead, we should be criticizing what is deeming the conclusion of instance arguments absurd in the first place. We should be questioning our belief that (EC) expresses a claim that is necessarily true on anyone's view.

Sawyer attempts to answer this question by doing three things. First she offers an explanation for why it is that instance arguments can systematically yield truths about our environment when used correctly. Next she identifies a causal feature that is usually not present when we make what we consider to be invalid inferences from introspective knowledge to empirical knowledge. It is the lack of this causal feature that elicits our doubts in the acceptability of instance arguments. Sawyer aims to show that instance arguments aren't

inherently unacceptable because when this causal feature is present it justifies the inference from introspective to empirical knowledge. Once this has been proven the last factor of Sawyer's argument is that it is semantic externalism which necessitates that this causal feature will always be present in instance arguments. As long as this feature is present in the instance arguments the inference from introspective to empirical knowledge will be acceptable. (Sawyer 529)

Sawyer explains the reliability of instance arguments through comparing concept-acquisition to photography. Photography is a process through which we take in information about the current state of the external environment by imprinting light reflections on paper so that we can use the information about that given state of the environment at some later time. Although the result of this process is simple, the process itself is rather complicated. Similarly, concept-acquisition is a process of storing information about the current state of the external environment upon its acquisition precisely because for any given concept there is 'one of a unique set of possible causal processes' necessary for its acquisition. (Sawyer 529) According to Sawyer, the Argument from Privileged Access relies on the similarities between photography and concept acquisition. She refers to concepts as resembling 'mental photographs' which can be used in our thought and communication at a later time and give us the ability to use them as evidence to determine the existence of some past state of the environment. (Sawyer 529)

The explanation of the reliability of instance arguments comes out when we show where the analogy between photography and concept-acquisition breaks down. Sawyer admits while it may not currently be physically possible, it is at least logically possible that someone could produce a fake photograph which depicts some information about a past state of the environment which did not occur. The reason this is important is because she argues that a similar situation with concept-acquisition could never happen because semantic externalism ensures that there is



no way that a particular concept could be acquired through a process that is not part of the unique set of possible causal processes necessary for its acquisition. Fake photographs are possible because someone could potentially create the exact right amount and kinds of light in just the right places but due to semantic externalism there is no way to get a given concept without the exact right conditions necessary to acquire it. Therefore, Sawyer illustrates that semantic externalism necessitates a causal feature of instance arguments which is that, “*causal contact (either to a natural kind or to a linguistic community) is a necessary condition for the acquisition of a concept,*” and to make any claim that states otherwise is to deny semantic externalism which is on par with accepting incompatibilism. (Sawyer 530) In order to prevent possible objections to this claim, Sawyer considers the question of whether it is possible to be mistaken in our thought’s that we have a given concept? She offers the contrapositive of (PA) to show why this question would be mistaken:

(PA’) For all **x**, if **x** does not have the concept F, where concept F is an essential component of the thought that p, **x** does not think that she thinks that p.

Before Sawyer moves on to why semantic externalism guarantees the presence of this causal feature in instance arguments she believes that she must illustrate two disanalogies between inferences from photographs and inferences from concepts. In photography, inferences made from photographs will yield specific facts about the external environment but inferences from concepts will yield *general* facts about the external environment. (Sawyer 530) Sawyer believes this is one possible explanation for why it is possible to produce fake photographs but it is not possible to acquire ‘fake concepts’. If some form of causal contact is necessary in order to possess a certain concept there is no way to fake that causal contact but the correct amount and kinds of light necessary to produce a given photograph can be faked. The second difference is

simply that the main purpose of photographs is to store information about the past where the main purpose of concepts is to be used in thought and communication. However, this is no reason to think of concepts in a different manner since the reason we are able to use them for such purposes is because they do store information about the external environment. (Sawyer 530)

Here is where Sawyer presents a crucial feature of her argument which has been working beneath the surface until now. She has been describing instance arguments up to this point in the way that causes most people to think that they are unacceptable in the first place. Usually we take instance arguments as a way to go from introspective knowledge to empirical knowledge of the *current* state of the environment. It could easily be argued that any empirical knowledge of the current state of the environment is impossible regardless of resulting from introspective knowledge. Even with empirical investigation, the second you gain empirical knowledge of what was the current state of the environment that state becomes a past state of the environment. Therefore, Sawyer makes the qualification that the true purpose of instance arguments is not to give us knowledge of facts of one's current external environment but instead to give us knowledge of facts of the *past* states of our environment. (Sawyer 530) If we imagine Susan in a Twin Earth switching scenario where her concepts are replaced once they spend a sufficient amount of time in the different environment, it is possible for her to have thoughts about a water concept, rather than a twater concept, in a twater environment as long as it is before her concepts are replaced. If Susan were to use instance arguments to conclude that she was in an environment containing water during this period of time then she would be wrong in her conclusion. However, Susan could make a correct conclusion that in the recent past she inhabited an environment containing water.

Finally, we move on to Sawyer's argument that semantic externalism guarantees that this causal feature she described earlier is present in instance arguments. According to Sawyer, individualism is the theory that claims that causal contact isn't a necessity in order to acquire a concept. Instead causal contact is only required for knowledge that our concept has some referent in the external environment. Under individualism it is possible to imagine a person who could make inferences from their concepts to facts about the external environment prior to any actual experience of the environment. Essentially, individualism embodies the concern that most people generally have with instance arguments; they present a possibility of someone having knowledge of contingent facts about the environment, facts commonly construed as only being ascertainable empirically, prior to any causal contact with the environment. Semantic externalism, on the other hand, expresses claims which are the negation of the claim of individualism expressed above. Knowledge of contingent facts prior to causal contact is impossible under semantic externalism which means that the feature of instance arguments that elicits most people's concerns cannot be present in such arguments. Now this does not mean that any conclusions of instance arguments are acceptable because we still cannot gain knowledge from them about the current state of the environment. However, the fact that prior causal contact is necessary to possess a concept means that our ability to think a thought using that concept allows us to gain knowledge of the past state of our environment since in the past we must have had a certain kind of causal contact with that past state of the environment.

"Semantic externalism is precisely the view that an individual cannot be regarded as complete with the concepts of external contingents independently of any prior causal contact with a specific given environment. The self can no longer be regarded as an entity completely separate from her environment. As a result, the apparent clear divide between the mind and the

world is eroded. The world we inhabit determines our mental capacities, our ability to think certain thoughts.” (Sawyer 531-532) This essential tie between our mind and our environment ensures that our concepts are proof in themselves that the necessary empirical conditions required to acquire them were satisfied at some earlier point in time. Moreover, our privileged access to these concepts ensures that we know this fact. The problem that most people held with instance arguments were that they seemed to allow us to come to know contingent facts about the environment without delving into experience at all but this line of thought is mistaken. We did have to delve into experience to gain this knowledge; it is just that we didn’t have to go into experience in between the occurrence of the thought and the acquisition of the knowledge. The causal contact necessary for the acquisition of the concepts used in the thought occurred prior to the instantiation of the thought. Without that prior causal contact we could not have had that thought in the first place.

Sawyer concludes that through the conjunction of (PA) and semantic externalism we are able to have direct knowledge of the concepts we possess as well as of the necessary conditions for acquiring those concepts. She wants to illustrate that knowledge of the latter is only possible under the assumption that semantic externalism can be known, but for the moment it is fine that it is possible to be known. (Sawyer 532) Sawyer does not believe that her theory presents a problem for epistemology or for the philosophy of mind because in no way does her theory claim that we can come to know everything empirical a priori. We can only come to know empirical things through introspection that have already been acquired empirically. (Sawyer 532) All Sawyer’s theory states is that privileged access gives us infallible access to our concepts so if we *could* know the necessary conditions for their acquisition then we *could* know that in a past state of our environment those conditions were satisfied.

## Chapter X

This journey through the compatibility debate between Content Externalism and Privileged Access has gone back and forth between those who think they have found a solution and others who are convinced that there is no possible solution. Once Paul Boghossian arrived at the scene all the potential solutions had to get much more intricate because we could no longer obtain our self-knowledge through inference or introspection without completely refuting his arguments which appear to be ironclad. Some people like Kevin Falvey, Joseph Owens, Andre Gallois, and Maria Lasonen-Aarnio tried to side-step Boghossian's arguments and in some ways they could be said to have succeeded. Although even those who believe they succeeded do so with a grain of salt because they know that our concepts of externalism and/or privileged access were weakened in the process. It wasn't until Sarah Sawyer decided to challenge the absurdness of the possibility of a priori knowledge of the external world that a light could be seen at the end of the tunnel. As with everything in philosophy, not everyone would be convinced that Sawyer actually solved the problem. The purpose of this project was to go through the steps of the compatibility debate to see why it is in the current position that it is in today. In this conclusion I will argue that Sawyer's solution is the best solution for compatibility put forth so far and I will explain why it overcomes the McKinsey argument while meeting the restrictions of a priori self-knowledge set in place by Boghossian.

Sawyer's theory is essentially that given semantic externalism in order to acquire a concept there must have been prior causal contact with either a natural kind or a linguistic community. So unless we are willing to give up semantic externalism, whenever we have an ability to think a thought utilizing a particular concept we can know that at some point in time we

had causal contact with either the natural kind that the concept refers to or a linguistic community that already possesses the concept. The reason that this does not interfere with the way in which we can know our own thoughts is because we cannot come to know contingent facts about the *current* state of our external environment but instead only about *past* states of the environment. While we did not acquire a priori knowledge of the concepts used in the thought since it was through empirical investigation that we acquired them, we did access the content of the thought completely outside of experience. Therefore any knowledge that results from our access to that content was obtained outside of experience and thus a priori. This knowledge would also be authoritative because no one else has access to these thoughts in or out of experience so no one else could be in a better position to claim that we are mistaken. Authority doesn't mean infallibility so the possibility of error is still present but all that matters is that no one is in a better position than we are to discover potential errors.

Some believe that knowledge of this kind may not be considered direct because going from knowledge of individual concepts to how those concepts function together in a thought require processes such as inference and observation. It is hard to object to this argument since any mediating steps between a thoughts occurrence and the acquisition of knowledge could easily be said to do away with any attribution of directness. However, there are a few things to keep in mind; the first is that this problem began with a priori knowledge and it was later on that some author started to discuss things like direct and authoritative knowledge. So the lack of this knowledge being direct doesn't clash with our intuitions about our privileged access to our thoughts that started this debate in the first place. The second thing to keep in mind is that there are a number of ways that we can describe how we go from the occurrence of a thought to the acquisition of knowledge without using inference, observation, or any other mediating steps. One

such possibility is that whatever was going on subconsciously or in the external world that caused an occurrence of a thought could have prompted a gathering of relevant information from our concepts to construct the thought. We simply accessed the information that our subconscious mind already gathered and didn't have to infer the content of the thought from the concepts contained within it. Basically what I am arguing is that if one were to object to Sawyer's argument based on a lack of directness then we would have to consider it as a completely separate issue first because there are a number of different possibilities to account for.

The last part of Sawyer's argument is meant to show that as long as we hold onto semantic externalism we can be sure that any instance arguments we use will have the causal feature that makes them reliable. In doing so she also provides an argument for why we cannot give up semantic externalism by showing that the only other option is internalism. To accept internalism is to allow for the possibility of a priori knowledge of the external environment because under internalism our concepts do not necessarily depend on relations to their referents in the external environment. This is on par with accepting the possibility presented in the conclusion of the McKinsey argument which most people consider absurd. As long as you believe that we cannot have a priori knowledge of contingent facts of our external environment regardless of whether they are facts about its past or current state, you cannot subscribe to internalism. Assuming then that semantic externalism is adopted, our ability to think a thought with a given concept means that prior causal contact must have occurred since we could not possess the concept without it. However, a priori knowledge of contingent facts of past states of the environment should not concern us. The features which categorize the knowledge as 'a priori' do not imply that we acquired all the relevant information that led to the acquisition of the knowledge outside of experience. Clearly to have knowledge of anything in the external

environment one must experience the external environment through empirical investigation. The a priori feature just refers to our *access* to the content of the thoughts using the knowledge obtained in experience through empirical investigation. Furthermore, the a priori access is only to the self-knowledge that we are thinking a thought with a given content that was acquired from experience.

Now there are 2 reasons why Sawyer's theory is so far the best solution presented to the compatibility debate and therefore why I believe it solves the problem altogether. The first reason is because it allows for consistency in McKinsey's triad of propositions without changing our notions of semantic externalism or privileged access while also addressing the conditions set in place by Boghossian. If we recall, McKinsey's triad is:

- (5) Oscar knows *a priori* that he is thinking that water is wet.
- (6) The proposition that Oscar is thinking that water is wet necessarily depends upon E.
- (7) The proposition E cannot be known *a priori*, but only by empirical investigation.

Under Sawyer's theory, (1) is straightforward and E in (2) is simply the causal feature of semantic externalism that prior causal contact is required for Oscar's possession of his concept of water. On the other hand, (3) is not as simple. If semantic externalism can be known a priori then (3) is wrong because Oscar could then potentially know E a priori. If semantic externalism cannot be known a priori then E also cannot be known a priori. However, as we saw in earlier chapters some have argued that if semantic externalism cannot be known a priori then there is no longer a problem. In such cases we lose the potential ability to use instance arguments to generate knowledge of the external environment and the absurd conclusion which started all of this no longer exists. So for the sake of argument we will assume that semantic externalism is knowable a priori. Through a priori knowledge of his thought and semantic externalism, E would



be that Oscar has had prior causal contact with water or a linguistic community with the concept of water. The only possible instance where McKinsey could allow that E is knowable a priori is where he includes descriptivist concepts but in such a case the problem dissipates because Oscar's thought no longer depends on E. Knowledge of his thought does not lead to an absurd conclusion anymore but McKinsey would probably not consider it 'knowledge' anyways because it does not pertain to anything that exists. If McKinsey is to maintain his position that the triad is inconsistent then he only has one option which is to argue that E isn't knowable a priori and he claims that Oscar cannot know his thought a priori due to that fact.

Sawyer's theory addresses this problem. The only possible reason McKinsey could have for why semantic externalism is knowable a priori but E isn't is because we cannot know a priori whether our concept is a natural kind or descriptivist concept. For McKinsey, such knowledge should only be obtainable through empirical investigation. He never discusses whether or not semantic externalism is knowable a priori but since we are assuming it is in this hypothetical situation then it seems like this is where McKinsey's objection would most likely exist. Either way it does not matter because Sawyer's theory is not burdened by this interpretation of E.

In most cases McKinsey is right that we should not be able to know that water exists in the external environment as a natural kind or as a concept in a linguistic community because of the possibility of descriptivist concepts. Although in the case of Sawyer's theory, we have infallible access to our concepts and our memory is not restricted. We can remember if we obtained our concept of water from interacting with the substance, learning of it from other people who have the concept, or if we simply had a thought about a colorless, tasteless, and odorless liquid and developed a concept to represent this hypothetical substance. So in this

theory we can know the status of our concepts a priori since we do not need to do any empirical investigation for the information.

In previous chapters when we discussed this problem we were restricting memory because it seemed counterintuitive to the idea of a priori knowledge since knowledge of how we acquired a concept was obtained through our experience of the external environment. Sawyer's theory has it that we have a priori access to the content of our concepts which we have previously acquired and how we initially acquired a concept would definitely be the type of information stored in it. The only catch here is that we can only have a priori knowledge of the status of our concepts as they pertain to a past state of our environment. A scenario is consistent with Sawyer's theory where after our acquisition of a concept through causal contact with an object in the external environment all of the objects that the concept refers to are wiped off the face of the Earth. It is up for debate as to whether such concepts would become descriptivist concepts or if they would still be natural kind concepts despite the lack of existing referents. However, in such a case we clearly could not be said to have a priori knowledge of the current state of our external environment. We should view our knowledge of the status of all of our concepts in the same way because upon the occurrence of any given thought our knowledge of the status of the concepts used within it is limited to our last interactions with the referents of those concepts.

Overcoming McKinsey's triad of propositions with Sawyer's theory is a much more difficult task than using it to explain away the absurdity in the conclusion of Aarnio's argument she extracted from McKinsey's triad:

- (1) Suzy can know a priori that she is thinking that *p*.
- (2) The proposition that Suzy thinking that *p* logically implies the proposition that *E*.

Therefore,

(3) Suzy can know a priori that *E*.

All one must do is let *E* represent the proposition that Suzy at some time prior to thinking that *p* has had the prior causal contact required by semantic externalism to possess the concepts utilized in that *p*. A priori knowledge of semantic externalism is enough to allow for that possibility even without the infallible access to our concepts since we do not have to deal with any proposition claiming that *E* isn't knowable a priori. For the same reason overcoming the triad is also more difficult than ridding the absurdity of the conclusion in Andre Gallois's McKinsey\* argument:

(1)  $\Diamond (\exists x) (x \text{ knows a priori that } x \text{ thinks that } x \text{ lives in a watery world})$

(2) ETC

(3)  $\text{ETC} \rightarrow \Diamond (\exists x) (x \text{ knows a priori that ETC})$

(4)  $\Diamond (\exists x) (x \text{ knows a priori that ETC})$

(5)  $\text{ETC} \rightarrow \Box (x) (x \text{ thinks } x \text{ lives in a watery world} \rightarrow x \text{ lives in a watery world})$

(6)  $\Diamond (\exists x) (x \text{ knows a priori that (5)})$

(7)  $\Diamond (\exists x) (x \text{ knows a priori that } \Box (y) (y \text{ thinks that } y \text{ lives in a watery world} \rightarrow y \text{ lives in a watery world}))$

(8)  $\Diamond (\exists x) (x \text{ knows a priori that } x \text{ lives in a watery world})$

ETC, as we recall from Chapter 7, refers to externalist theories of content (content externalism). Gallois's solution was that our lack of an ability to know a priori whether our concepts are natural kind or descriptivist is enough for us not to be able to know a priori that we live in a watery world. While this solves the problem we lose the ability to know that the a priori knowledge we obtain from our own thoughts actually applies to anything in the external environment prior to doing empirical investigation. It seems wrong to say that when Oscar has the thought 'water is wet' he can't know if there actually is any such thing that his water concept

refers to in the world he inhabits until he encounters it *again*. I stress the word ‘again’ because Oscar has obviously encountered water before in one way or another if he has the concept, is able to use it in a thought, and knows it includes the property of wetness. On the other hand, Sawyer’s theory just assumes Oscar has had interactions with water before since his possession of the water concept requires it through semantic externalism. The absurdity of the conclusion in the McKinsey\* argument is removed through her theory by substituting the consequent of the second conditional in (5), ‘x lives in a watery world’, with ‘x lived in a watery world at some time prior to the occurrence of x’s thought that x lives in a watery world’. With this substitution the conclusion (8) would become  $\Diamond (\exists x) (x \text{ knows a priori that } x \text{ lived in a watery world at some time prior to the occurrence of } x\text{'s thought that } x \text{ lives in a watery world})$ . Again there is no need for infallible access to concepts because (6) essentially states that it’s possible for someone to know a priori the equivalent of proposition E.

The real challenge for Sawyer’s theory does not lie in the different versions of the problem presented by McKinsey. Boghossian was the one who brought the real challenge to compatibilists. He believed that semantic externalism and privileged access had to be compatible because he truly felt that we do have a special access to our mental states despite the hold that semantic externalism has on them. Though he also believed that no argument could think of or had read from others would actually solve the problem. The reason that no argument would suffice was because in each argument the self-knowledge they entailed had at least one of three features that a priori self-knowledge could not exhibit. In order to show that Sawyer’s theory has the potential to be a true solution to the problem it must be tested against these three features to see whether the self-knowledge it allows for exhibits any of them. According to Boghossian, a priori self-knowledge cannot be acquired through inference or inner observation. In the same

regard it also cannot be acquired from or based on nothing. As we have seen so far in Sawyer's theory, if the resulting a priori self-knowledge cannot be argued to be the result of inference or inner observation then it almost certainly could be argued to be based on nothing. However, it is the way in which Sawyer's theory works that it disagrees with Boghossian's whole argument for why a priori self-knowledge cannot exhibit these features in the first place.

Boghossian's main argument for why inference cannot be used to acquire self-knowledge is because it will always result in an infinite regress in one way or another. When discussing what justifies our a priori knowledge of an occurring mental state, the justification will have to rest on another mental state since the a priori factor restricts us from using anything outside of our mind. The mental state on which the justification rests will itself have to be justified by yet another mental state and so on. A different kind of infinite regress ensues when we do not require that our knowledge of every mental state involved must itself be justified in order to have justified knowledge of an occurring mental state. Perhaps all that we need to have justified knowledge of an occurring mental state is that this knowledge is justified by knowledge of another mental state. The key here is that the knowledge of this latter mental state does not have to be justified knowledge as long as the knowledge is sufficient to justify the knowledge of the occurring mental state. This would appear to solve the problem but this is just an illusion because the acquisition of the latter knowledge must have gone through the same process when it was an occurring mental state. So in order to have knowledge of that state we must have knowledge of yet another state. Instead of an infinite regress of justification we have an infinite regress of knowledge of particular mental states that must be known before we can have justified knowledge of an occurring mental state.

In Sawyer's theory she describes the process through which we obtain our self-knowledge of our thoughts. She claims that we obtain knowledge of our thoughts through introspection and through conceptual analysis we obtain knowledge of the fact that semantic externalism ensures that prior causal contact must have occurred for us to possess the concepts used in the thought. If we make the jump to knowledge that our concepts must have referents in the external environment then that knowledge was acquired through inference from the first two bits of knowledge. The knowledge resulting from introspection is something that will be discussed when we get to Boghossian's second section. As for conceptual analysis, it should be an unproblematic method of obtaining self-knowledge for Boghossian because if you can have a priori knowledge of semantic externalism then its implications should also be knowable a priori. On the other hand, the knowledge resulting from inference that we must have had prior causal contact to possess the concepts in our thoughts does run into problems with Boghossian's first section.

At first glance, Sawyer's theory could be in a bit of a problem because it involves self-knowledge resulting from an inference. The knowledge that we live in a world where referents exist for the concepts we used in our thought is justified by combining the knowledge of our occurring mental state with our knowledge of semantic externalism. An infinite regress appears to unfold because the knowledge of our occurring mental state must be justified by our knowledge of some prior mental state. Within Sawyer's theory this latter mental state which does the justifying would be the mental state that occurred when we had the necessary prior causal contact with those referents of the concepts used in the occurring mental state. The knowledge of the latter state would then be justified by knowing that in the past our knowledge of our mental states has reliably allowed us to come to true judgments about the external world. However, in

Sawyer's theory this regress does not actually have to occur because the knowledge that we possess a given concept is enough to justify the knowledge which results from the inference. Our ability to think a thought using the concept ensures that we possess the concept and knowledge of semantic externalism allows us to know that our possession of the concept guarantees that we acquired it through the required kind of contact with its referents. While knowledge of other mental states might help provide further justification, this additional justification is not necessary. Sawyer's claim that we have infallible access to our concepts is what ensures that we do not have to worry about whether or not our concept is a natural kind or descriptivist concept. As long as we can remember the manner in which we acquired the concept we can know whether we obtained it through causal contact or if we hypothesized about it.

As long as the only inference used in this theory is the inference from knowledge of externalism to knowledge that prior causal contact is necessary for the possession of a concept then it does not clash with the direct feature of privileged access. Once this inference is made then it becomes knowledge that we can use whenever we wish to. If we already have the knowledge before the occurrence of a thought then we can simply know without the use of any further inferences that our possession of the concepts used in the thought requires that we had prior causal contact with certain objects in the external environment.

If someone were to object to this argument by claiming that more inferences must occur in between the occurrence of a thought and obtaining a priori knowledge of its content then it would present a problem for how the theory accounts for the directness feature. The first point to be made is that the inference described above only pertains to how we acquire a priori knowledge of the status of our concepts used in our thoughts. If an inference is necessary in order to simply know the content of our thoughts then this would be a problem but there does not

appear to be any such inferences made for the acquisition of that knowledge. Inferences would be required to obtain both the a priori knowledge of the status of our concepts as well as the conclusions of instance arguments but the simple observation that leads to the a priori knowledge of the content of a thought does not use any inferences whatsoever. Any objection by this route would need to show that inferences must have occurred in order to acquire the a priori knowledge of the content of a thought but at this point in time I have neither seen nor could I imagine any such arguments.

In Boghossian's second section he argues that self-knowledge cannot be the result of inner observation. According to Boghossian, the inspection of an object can at most provide us with knowledge of its intrinsic properties. Things like extrinsic or relational properties can only be discovered through empirical investigation. He maintains that in order to us to be able to say we know an object we must be able to distinguish that object from other objects which could be considered relevant alternatives. In the same regard, the introspection of our own thoughts can at most provide us with knowledge of the intrinsic properties of our thought's as well as the concepts within them but the ability to distinguish those concepts from other concepts which could be considered 'counterfeits' requires that we have knowledge of the thought's and/or concept's extrinsic or relational properties. Therefore he concludes that introspection, or inner observation, cannot be the process through which we acquire self-knowledge of our mental states.

In Sawyer's theory, the knowledge we obtain of our occurring mental states is acquired through introspection. Boghossian would argue that we cannot distinguish this mental state from other mental states which contain at least one different 'counterfeit' concept. However, Sawyer's notion that we have infallible access to our concepts does allow us a way around this problem.



Since we can remember the event in our life when we acquired the concept used in the thought we can know exactly what concept it is. Although it is not as easy to explain away the problem presented in Boghossian's Twin Earth thought experiment where the subject is transported back and forth between Earth and Twin Earth without their knowledge. Here our subject knows that they have had prior causal contact with the substance they call water but they are unaware of the fact that they've actually come in to contact with two distinct substances called water. All of the technology available to the subject is insufficient for discovering the difference between the substances.

In this scenario our subject cannot distinguish whether their thought about water was actually about water or 'twater'. Sawyer does not address this thought experiment but I believe that her theory can still account for it. The subject can know that they have had prior causal contact since they know they possess the water concept but they just don't know that they've been in prior causal contact with two distinct substances. So instead of the subject possessing two distinct concepts that they can't differentiate from each other they actually just have two distinct substances labeled under one concept since to the subject both substances have identical properties. It makes no sense to claim that the subject's unconscious mind can identify the difference between the substances and classify them under two distinct water concepts when the subject consciously has no ability to do so. The most likely situation is that both substances with fall under the same concept until the subject is able to discover the difference between them and once discovered they will create a new concept so the two substances can be properly classified. So in this thought experiment, it is true that the subject cannot tell the difference between their water and 'twater' thoughts. However, for Sawyer's theory there aren't actually going to be

distinct water and 'twater' thoughts because both substances have identical properties to the subject so they both fall under one water concept.

Essentially Sawyer's theory would address Boghossian's problem with inner observation by claiming that if a subject observes a difference in properties between two substances then they will be able to classify them under distinct concepts and they will be able to distinguish between thoughts using both concepts. This is because the information that pertains to the differences between them will be included in the concept and the subject has infallible access to this information when their mental state using either concept occurs. However, if a subject is not able to observe any difference between the properties of two distinct substances then they will naturally classify them under the same concept. Properties are what we use to differentiate one natural kind from another so if a subject observes exactly the same properties in two substances they have no reason to think that the two substance do not fall into the same category.

One objection to this argument is that it doesn't address Boghossian's argument that inner observation at most yields knowledge of intrinsic properties and it is instead the extrinsic or relational properties that determine the content of a thought. This only matters when we were assuming the only information available to our inner observation must be obtained outside of experience. In Sawyer's theory all of our interactions with a substance in experience are stored as information in our concept of that substance and this information is available to us when we use inner observation to acquire the knowledge of the content of our thoughts. While inner observation alone isn't enough to yield knowledge of extrinsic properties, when you combine it with infallible access to the information stored in our concepts then it can actually allow for knowledge of extrinsic properties. As long as there are methods to obtain knowledge of extrinsic properties then we can use those methods in experience, store the information in the concepts,

and retrieve the information when using inner observation on our thoughts. So in Sawyer's theory she actually agrees with Boghossian that observation alone can only yield knowledge of the intrinsic properties but this misses the idea that empirical investigation can give us knowledge of extrinsic properties which can be stored and retrieved later through inner observation.

In Boghossian's last section he argues that self-knowledge cannot be based on nothing; or in other words it must be considered a cognitive achievement to be considered knowledge. The main reason for the inclusion of this section is because Boghossian could not imagine what else could justify self-knowledge other than inference or inner observation so the last option to consider is self-knowledge based on nothing. In the previous paragraphs I argued that self-knowledge in Sawyer's theory is a result from inference and inner observation but it does not create the same problems that Boghossian thought they did. Since this self-knowledge is a result from inference and inner observation it will necessarily be classified as a cognitive achievement. In this section Boghossian argues that self-knowledge which doesn't result from inference or inner observation is still a cognitive achievement because it doesn't exhibit three characteristics that self-knowledge which is not a cognitive achievement would exhibit. As long as it is a cognitive achievement it cannot be based on nothing but this does not present a problem for Sawyer's theory because having self-knowledge not be considered a cognitive achievement is sort of a last resort for compatibilists. Most philosophers would not consider knowledge which is based on nothing to be knowledge in the first place because it does not actually apply to anything. So the fact that Sawyer's theory does not address this section is actually a good thing because it shows that the self-knowledge that it yields is considered a cognitive achievement and is applicable.

It is very well possible that sometime in the future there will be more incompatibilist arguments developed that compatibilist theories must address before they can be truly considered a potential solution. At that point in time Sawyer's theory could be once again called into question. However, at this point in time the McKinsey argument and Boghossian's arguments are the strongest incompatibilist arguments available and Sawyer's theory is able to address them both. So far it is the only theory which can account for the potential problems created from combining content externalism and privileged access while also maintaining consistency with our intuitions about both. It feels safe to say that until new incompatibilist arguments are developed, Sawyer's theory is the best fallback for compatibilists to justify how they can know the content of their thoughts while living in this world ruled by externalism.

## Works Cited

1. Boghossian, Paul. "Content and Self-Knowledge." *Externalism and Self-Knowledge*. Ed. Peter Ludlow and Norah Martin. Stanford, Calif: CSLI Publications, 1998. 149-173. Print.
2. Brueckner, Anthony. "What An Anti-Individualist Knows A Priori." *Externalism and Self-Knowledge*. Ed. Peter Ludlow and Norah Martin. Stanford, Calif: CSLI Publications, 1998. 197-205. Print.
3. Burge, Tyler. "Individualism and Self-Knowledge." *Externalism and Self-Knowledge*. Ed. Peter Ludlow and Norah Martin. Stanford, Calif: CSLI Publications, 1998. 111-127. Print.
4. Davidson, Donald. "Knowing One's Own Mind." *Externalism and Self-Knowledge*. Ed. Peter Ludlow and Norah Martin. Stanford, Calif: CSLI Publications, 1998. 87-110. Print.
5. Falvey, Kevin, and Joseph Owens. "Externalism, Self-Knowledge, and Skepticism." *The Philosophical Review* 103.1 (1994): 107-37. Print.
6. Gallois, Andre. *The World Without, The Mind Within*. New York: Cambridge University Press, 1996. Print.
7. Lasonen-Aarnio, Maria. "Externalism and A Priori Knowledge of the World: Why Privileged Access Is Not the Issue." *Dialectica* 60.4 (2006): 433-45. Print.
8. McKinsey, Michael. "Anti-Individualism and Privileged Access." *Externalism and Self-Knowledge*. Ed. Peter Ludlow and Norah Martin. Stanford, Calif: CSLI Publications, 1998. 175-184. Print.
9. Sawyer, Sarah. "Privileged Access to the World." *Australasian Journal of Philosophy* 76.4 (1998): 523-33. Print.

## Vita

Kevin Sadekoski was born in Utica, New York. After graduating from Christian Brothers Academy, Colonie, New York, in 2005, he entered into his Undergraduate work at Syracuse University in Syracuse, New York. He received a Bachelor of Art degree double-majoring in Philosophy and Psychology from Syracuse University in May 2009. In August 2010, he started his Graduate work pursuing a Master of Art degree in Philosophy at Syracuse University.